New Diagnostic Tools for Capture-Recapture Models

Anita Jeyam

A thesis submitted for the degree of
Doctor of Philosophy in the subject of Statistics

School of Mathematics, Statistics and Actuarial Science

University of Kent

October 2017

Abstract

Capture-recapture models have increased in complexity over the last decades and goodness-of-fit assessment is crucial to ensure that considered models provide an adequate fit to the data. In this thesis, my primary emphasis is to develop new diagnostic tools for capture-recapture models in order to target possible reasons of lack-of-fit, which might provide biological insights and point towards better-fitting candidate models.

Starting with the basic Cormack-Jolly-Seber model, I develop a new tool for detecting heterogeneity in capture. I then progress to the more complex multi-state models, for which I propose a test for detecting a mover-stayer structure within the population. Finally, I move on to more general models presenting additional levels of uncertainty: first partial observations and then unobservable states. In the presence of partial observations, part of the observations are assigned to states with certainty whereas others are not. I develop a new test for the underlying state-structure of the partial observations, this test detects that the partial observations are not generated by the observable states defined in the experiment. In the presence of unobservable states, the additional level of uncertainty relates only to the non-captures. I present a procedure to test whether one or two unobservable states need to be defined in order for the model to provide an adequate fit to the data.

Lastly, I explore the use of multi-state models to incorporate individual time-varying covariates, based on a fine discretisation of the covariate space.

Acknowledgements

I would like to thank my supervisor Dr Rachel S. McCrea for being so supportive, patient and encouraging throughout the PhD, and for making it such a great experience. I am also very grateful to Rachel for going through so many reports and draft versions, and in general for her advice, for being available and making time for me. I would also like to thank my other supervisor, Dr Roger Pradel for his support and kindness, for making visits to the University of Kent and having lengthy meetings or Skype calls about the PhD work. Finally, I would like to thank the other members of my supervisory team: Dr Diana J. Cole for explaining the concept of parameter redundancy, for providing and explaining the use of her Maple functions, and going through chapter drafts; as well as Prof. Byron J.T. Morgan for all his helpful comments on the thesis. I would also like to thank Byron for his guidance and all the discussions we had prior to starting the PhD.

I would like to thank the CEFE in Montpellier for welcoming me during my first year. I thank Dr R. Choquet for providing Matlab code used for the goodness-of-fit tests implemented in program U-CARE, and Dr O. Gimenez for his helpful comments regarding the work from Chapter 2. I would also like to thank both the CEFE and my supervisory team as a whole for enabling me to attend the E-SURGE workshop in Montpellier.

I also would like to thank everybody at the Tour du Valat for welcoming me there and taking me out on so many field outings (thanks to the supervisory

team for making this possible in the first place). I would like to thank Prof. R. Griffiths for letting me join with the Great crested newt fieldwork at the University of Kent.

I thank Dr T. Bregnballe and Dr M. Frederiksen for their ecological insights and comments on the Chapter 2 work, and for providing the Sandwich tern dataset. Thank you also to Dr R. Langrock for providing the HMM R code for model fitting and simulations for Chapter 6, and thanks to Dr J. Nichols for providing the Meadow vole dataset for Chapter 6.

I thank the School of Mathematics, Statistics, and Actuarial Science (SMSAS) of the University of Kent and the National Centre for Statistical Ecology (NCSE) for funding this PhD and giving me this opportunity. I thank the whole SE@K group at Kent for being such a great team to be part of and for all the fun moments we had. I thank my PhD colleagues, Tita and Ela especially, for all our discussions and their company during this sometimes tough process. I would also like to thank Claire Carter and Judith Broom from SMSAS for always kindly and patiently helping out with my admin, paperwork etc. issues.

Finally, I would like to thank my mother and my father for providing me with so many great educational opportunities throughout my life and encouraging me to pursue them. I would also like to thank my friends for all their support. In particular, thank you Twinny for always having my back and thanks to Pompoen and Fairy for the motivational boosts and just being endlessly inspiring. Last but not least, I thank my aunt, Dr Siron Rajaratnam, for being my inspiration and role model throughout my life.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In the current context of global climate change, environmental issues are the object of many government policies. Measuring and understanding environmental change, and its impact is key to generating suitable and efficient policies. This is an incredibly vast puzzle, spanning several areas such as climate change modelling, pollution measurement, or more generally ecosystems' understanding; and studying the dynamics of animal populations is a piece of this puzzle.

1.1 Capture-mark-recapture

Unlike humans, for whom census is (generally) straightforward, specific data collection techniques have to be used to monitor animal populations. One of the techniques used to study a species consists of marking animals from a population of interest uniquely and then recording whether they are recaptured or re-sighted at subsequent sampling (capture) occasions, which are generally made at equal time intervals. The resulting data are known as capture-mark-recapture (CMR) data (Williams et al., 2002). Note that the vocabulary relating to capture, sighting, encounter will be used interchangeably throughout this thesis.

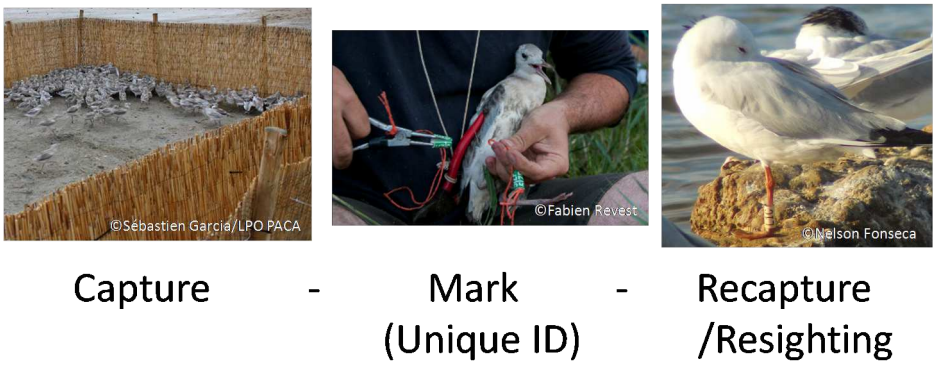


Figure 1.1: Capture-mark-recapture of slender-billed gulls.

Different kinds of markings are used depending on the species of interest: rings are commonly used for birds (e.g. flamingoes, gulls), ear tags for ibex, pen-markings for lizards, and so forth. Figure 1.1 illustrates the CMR technique conducted on slender-billed gulls (*Chroicocephalus genei*). Marking can be unnecessary for some species that already present natural individual marks, akin to human fingerprints: for instance, Great crested newts' (*Triturus cristatus*) belly pattern (McCrea and Morgan, 2014, p.3), or the dorsal fin and scars for various whale species (see for example Hammond et al., 1990).

The CMR technique is not restricted to ecological applications. It is widely used in other areas where populations are difficult to follow such as forced labour (van der Heijden et al., 2015) or drug use (Mastro et al., 1994).

1.2 Capture-recapture models' increasing complexity

The capture-recapture models used to understand the mechanisms underlying data collected from a CMR experiment have become increasingly complex over the years, in order to be more biologically realistic and thus result in more accurate inference. This thesis focusses on models aiming to estimate survival

and/or movement within an open population, allowing for birth/immigration and death/emigration (McCrea and Morgan, 2014, p.57).

The most basic capture-recapture model is the Cormack-Jolly-Seber (CJS) model, which estimates the apparent survival probability within the analysed population, whilst accounting for imperfect detection (Lebreton et al., 1992). The survival is termed as apparent because permanent emigration and death are confounded.

However, other demographic parameters are also of interest such as probabilities of movement between colonies for migratory birds, or transitions between physiological states like health status. Hence, multi-state models were introduced, to estimate state-dependent survival probabilities and the probability of moving between states, whilst accounting for imperfect detection in each of the states (McCrea and Morgan, 2014, p.87) . The process governing movement between the different states is a Markov chain, usually assumed to be of first-order; meaning that the future state depends on the past only through the current state. Or, in other words, there is no memory. This is not always realistic since animals such as Canada Geese have been shown to display memory (Hestbeck et al., 1991). The order of the Markov chain can be increased to account for this kind of issue, using for example a second-order Markov process (i.e. the future state depends on the past through both present and previous state) (see for example McCrea and Morgan, 2014, p.96) .

If the states of interest are animal locations (e.g. to study migration patterns), states are easily assigned during fieldwork: when an animal is captured, its state is clearly visible and recorded. However, when these states are physiological, like health status, inference often has to be made based strictly on the behaviour of the animals as it is often impossible to perform medical tests on them. In this case, the state assignment is prone to error and more complex models were developed to account for this uncertainty: multievent models (Pradel, 2005). They form a general umbrella model, encompassing the

multi-state and CJS models, which can both be easily expressed as multievent models.

Finally, the model parameters can be time-dependent and/or state-dependent, or constant overall; they can also be expressed as functions of covariates that may be relevant (e.g. gender, climate) (Lebreton et al., 1992).

Hence, the set of potential models which can be fitted to a single dataset is very large.

1.3 The necessity for diagnostic tools

'Essentially, all models are wrong, but some are useful' (Box and Draper, 1987).

Statistical models are, by definition, based on assumptions and subsequent inference is only valid under those assumptions. In other words, unrealistic assumptions can lead to flawed inferences. The increasing complexity from CJS to multievent models, is due to a relaxation of some of the more strict assumptions, in order to obtain more informative and accurate biological conclusions.

Many competing models are generally fitted to a dataset. In order to draw biological conclusions and make inference, a selection must be made. This is usually done using information criteria such as the AIC (Akaike's Information Criterion), or tests for nested models such as the likelihood-ratio test or the score test. In certain cases, it is not necessary to select only one model and model averaging may be performed. This procedure should however be used with caution as only compatible models should be averaged together (Newman et al., 2014, Chapter 5). At this stage, the best model(s) among the fitted models is chosen. The candidate models most often do not exhaustively represent the set of all possible models. Particularly for complex models, it would be too time-consuming to explore all possibilities, the likelihood optimisation

might also prove problematic for some models (McCrea and Morgan, 2011). Thus, the model selected by information criteria is the least worst among the investigated candidates, whilst the likelihood ratio test and score test also assess the relative fit of examined models. It is crucial that at least one of the candidate models should provide an adequate fit to the data (Pradel et al., 2003): if the selected best model does not fit the data adequately, erroneous conclusions may result, which in turn can be misused by decision-makers.

Assessing whether a model fits data adequately is termed absolute goodness-of-fit assessment. It is usually measured overall, based on the distance between expected values under the model and observed values. The statistic most commonly considered, for data grouped into K cells, is Pearson's χ^2 (Cressie and Read, 1988, page 10). Let O_k and E_k respectively denote the observed and expected frequencies in cell k. The test-statistic is then defined by:

$$\chi^2 = \sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k} \,. \tag{1.1}$$

Capture-recapture data can be viewed as multinomial data, with all the possible capture histories representing the outcomes, and the number of animals with each capture history following a multinomial distribution. Hence, one way of examining goodness-of-fit would be to compare expected and observed values for these frequencies. This general goodness-of-fit assessment is straightforward to understand and easily implemented, but not optimal for a capture-recapture framework. Indeed, the number of possible capture histories increases exponentially with the number of capture occasions: $(S+1)^{K-1} - 1$ histories for an experiment with K capture occasions and S states, and this, in turn, leads to very sparse data for goodness-of-fit purposes (Lebreton and Pradel, 2002; Pradel et al., 2005; Lebreton et al., 2009). In addition to this, Burnham (1991) also noted that the omnibus test has low power.

More importantly, even in perfect conditions, an overall goodness-of-fit test

does not provide enough information. Indeed, if the omnibus test is rejected, we can conclude that the model does not fit the data adequately, but there is no indication as to why it does not fit (Pradel et al., 2005). Diagnostic tests on the other hand, are specific to model assumptions, and thus provide information regarding possible reasons for lack-of-fit. This guides the model-building process by pointing towards more appropriate models for the analysed data and can also provide new ecological insight. Hence, diagnostic tools are particularly valuable for capture-recapture models.

Note that lack-of-fit may also result from over dispersion (Amstrup et al., 2005, p.19) (higher residual variance than expected under a multinomial model) and this is usually accounted for by using a variance inflation factor, the $\hat{c}$, which is the ratio of the chi-square statistic divided by the number of degrees of freedom (McCrea and Morgan, 2014, p.174).

1.4 Thesis structure

The main motivation of this thesis is to develop new diagnostic tools for capture-recapture models, that are relatively easy to implement and interpret. Their performance is generally evaluated using simulation and they are also applied to real-life datasets. This thesis explores potential reasons of structural lack-of-fit of the models and does not consider over dispersion assessment. Also, the aim of this thesis is to develop goodness-of-fit tools for capture-recapture models, and thus we do not present the technicalities of parameter estimation.

Diagnostic tools currently exist for the CJS model as well as for multi-state models (Pradel et al., 2005). They detect specific phenomena that cause a breakdown of the model assumptions such as trap effects and transience. These concepts will be explained in Chapter 2, alongside the CJS model and its existing diagnostic tests.

Other phenomena might also violate model assumptions and it would be of interest to be able to detect these and identify them specifically. Chapter 2 focusses on heterogeneity in capture within a CJS framework and introduces a test of positive association, based on Goodman-Kruskal's gamma (Siegel and Castellan Jr., 1988) as a new tool for detecting heterogeneity in capture.

Chapter 3 progresses to multi-state capture-recapture models and naturally extends the test of positive association to the area of transition probabilities, which have generally not been given much attention in the literature. Details are then provided for the only existing diagnostic test related to the transition probabilities: Test WBWA, currently used to detect memory. Finally, Chapter 3 gives rise to a new tool to detect the existence of a mover-stayer structure within the analysed population by combining the existing Test WBWA and the test of positive association.

There are currently no diagnostic tests for the more complex multievent capture-recapture models, which account for uncertainty in state assignment. We approach the area of uncertainty by examining two specific cases: partial observations in Chapter 4, and unobservable states in Chapter 5. Partial observations occur when the state cannot be determined upon capture for some animals (i.e. state uncertainty for some observations); whereas unobservable states present an additional level of uncertainty only on the non-captures. We developed tests of the model structure in terms of underlying states for both these cases; and our main aim was to assess whether they could work. Their performance was therefore assessed using simulation under extremely good conditions, i.e. very large sample sizes.

Chapter 4 presents the multievent framework and the existing mixture test, Test M, currently used for multi-state models. We show that a mixture test can also be defined for partial observations. This new test detects whether there is evidence that partial observations are actually generated by other states than those directly observable in the experiment.

Chapter 5 develops a procedure to test for unobservable states. In order to do so, it was necessary to examine a more general framework of mixtures of multinomials, which we present in detail. We were then faced with the obstacle of parameter redundancy, which is defined and explored in detail for our situation. We derive general parameter redundancy results for mixtures of multinomials in the presence of unobservable components. We finally implement a procedure for the general framework of mixtures of multinomials to test for unobservable components. We then apply this procedure within a capture-recapture framework, to test whether there is statistical support for defining unobservable states.

Chapter 6 constitutes a separate but related piece of work. We mentioned previously that covariates can be incorporated in capture-recapture models. However, this is particularly challenging for time-varying individual covariates such as weight, since on each occasion the animal is not captured, its weight value is unknown. Different methods have been proposed to handle this problem, and Chapter 6 explores the use of multi-state models with a large number of states representing discretised values of the covariate.

Finally Chapter 7 summarises the contributions of this thesis and further research paths that can be explored.

Chapter 2

A new tool for detecting heterogeneity in capture

This chapter proposes a new method to detect and identify heterogeneity in capture within a Cormack-Jolly-Seber (CJS) framework. In Section 2.1, we describe the CJS model and associated data, we also detail the existing diagnostic tests derived for this model. The CJS model is based on relatively restrictive assumptions, which are detailed in Section 2.1. This chapter focusses on the assumption of equal recapture probability at each occasion for all marked animals known to be in the population.

This assumption is violated when there is a trap-effect (i.e. capture at a given occasion affecting capture probability at the following occasion); but also more generally when the animals demonstrate heterogeneous behaviour in terms of capture. This is very common in animals with a social structure, where dominant animals are more likely to be resighted than subordinates (Cubaynes et al., 2010), or if some animals are located in places that are difficult to access.

It is important to identify and account for heterogeneity in capture when it occurs: firstly, not accounting for it can lead to biases in estimates of demo-

graphic parameters such as survival. Although survival estimates have been shown to be fairly robust, even small biases can lead to flawed inference or have an impact on management strategies (Prévot-Julliard et al., 1998; Cubaynes et al., 2010; Fletcher et al., 2012; Abadi et al., 2013). Secondly, the presence of heterogeneity in capture will warrant further investigations as to its causes, this may in turn lead to identifying individuals with different behavioural patterns such as breeders/non-breeders, bold/timid, dominant/subordinates or animals with different feeding strategies (Corkrey et al., 2012). Finally, heterogeneity in capture can also be a result of the study design (Oliver et al., 2011; Corkrey et al., 2012), and identifying it would give directions to possible adjustments. Indeed, heterogeneity in capture can be related to the sampling process, or stem from resighting errors, in which case additional data collection rules may be specified: for example, it is more or less standard practice to require at least two observations for neck-banded geese (Madsen et al., 2014).

It is currently difficult to identify heterogeneity in capture separately from other phenomena, particularly trap-dependence, that could cause a breakdown in the CJS model assumptions. In Section 2.2, we introduce a new tool to detect heterogeneity in capture, based on Goodman-Kruskal's gamma (Siegel and Castellan Jr., 1988). Other methods which have been proposed to identify this phenomenon are presented in Section 2.3: an adapted version of the existing diagnostic tests, suggested by Péron et al. (2010), Leslie's equal catchability test (Orians and Leslie, 1958) and Carothers' extension of the Leslie test (Carothers, 1971).

A simulation study is conducted in Section 2.4 to assess the performance of the new approach in detecting heterogeneity in capture specifically, comparatively to the other methods. In Section 2.5 we apply the tests to a dataset of Great cormorants (*Phalacrocorax carbo sinensis*) and a dataset of Sandwich terns (*Thalasseus sandvicensis*). Finally, we conclude and discuss the results obtained in Section 2.6.

2.1 The Cormack-Jolly-Seber framework

2.1.1 Data summaries

Capture-recapture data in its simplest form consists of records of whether each animal is seen (coded as 1) or not seen (coded as 0), at each sampling occasion i of a capture-recapture experiment (see for example Lebreton et al., 1992). This produces a set of encounter or capture histories, as illustrated in Table 2.1. In this example, animal *Z321* was first captured, marked and released at occasion 1, not seen from occasions 2 to 5 and was seen again at occasion K. The data can be compressed in a capture history matrix, by pooling animals with the same capture history, as shown in Table 2.2. The encounter histories can be further summarised by considering pairs of release and subsequent recapture: the full m-array, which classifies the animals by previous capture history, and the reduced m-array which conditions only on current capture (Burnham et al., 1987). All animals released at the same occasion form a cohort, and the animals from the same cohort, which share the same previous capture history, form subcohorts . Table 2.3 provides an example of a full m-array; it contains the number of animals released at occasion i and recaptured for the first time at occasion j, or never seen again, grouped by their previous capture history $\{h\}$. Table 2.4 presents a reduced m-array m, with elements $m_{i,j}$ and v_i defined as the animals released at occasion i and first recaptured at occasion j, and those never seen again, respectively.

Table 2.1: Example capture histories for a capture-recapture experiment with K encounter occasions

Animal ID	**Encounter occasion**						
	1	**2**	**3**	**4**	**5**	...	**K**
Z321	1	0	0	0	0	...	1
C324	1	1	0	1	0	...	1
K114	0	1	0	0	0	...	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
Z124	0	0	1	1	0	...	0

Table 2.2: An example of summarised capture history matrix with 6 encounter occasions

Capture history						**Number of animals**
1	1	1	1	1	1	10
1	0	1	0	1	0	15
1	0	0	0	0	1	22
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0	0	0	1	1	0	28
0	0	0	1	0	0	4
0	0	0	0	1	1	36

Table 2.3: Full m-array example: i denotes the release occasion, j denotes the first subsequent recapture occasion, and $\{h\}$ denotes the previous history of the animals

		j					Never seen again
		2	**3**	**4**	**5**	**6**	
$i\{h\}$	**1**{-}	10	15	0	0	22	0
	2 {1}	-	10	0	0	0	0
	{0}	-	0	0	0	0	0
	⋮	-	-	⋮	⋮	⋮	⋮
	5 {1 1 1 1}	-	-	-	-	10	0
	{1 0 1 0}	-	-	-	-	0	15
	⋮	-	-	⋮	⋮	⋮	⋮
	{0 0 0 1}	-	-	-	-	0	28
	{0 0 0 0}	-	-	-	-	36	0

Table 2.4: Reduced m-array: i denotes the release occasion, j denotes the first subsequent recapture occasion

	j					Never seen again
i	**2**	**3**	**4**	**5**	**6**	
1	19	20	3	0	22	0
2	-	19	11	35	0	0
3	-	-	47	29	6	0
4	-	-	-	50	49	7
5	-	-	-	-	46	104

2.1.2 The CJS model

Capture-recapture data consisting of 1s and 0s, as presented in Section 2.1.1, are typically modelled using a CJS model, which conditions on first capture. It allows the estimation of the apparent survival probability from occasion i to $i+1$, ϕ_i (the model does not distinguish death from permanent emigration), as well as the imperfect recapture probabilities at occasion i, denoted p_i (see for example McCrea and Morgan, 2014, p. 70).

As stated in Chapter 1, the CJS model is based on relatively restrictive assumptions. They are listed below, as defined in Williams et al. (2002).

1. Every marked animal present in the population at sampling period i has the same probability p_i of being recaptured or resighted.

2. Every marked animal present in the population immediately following the sampling period i has the same probability ϕ_i of surviving until sampling period $i+1$.

3. Marks are neither lost nor overlooked, and are recorded correctly.

4. Sampling periods are instantaneous.

5. All emigration from the sampling area is permanent.

6. The fate of each animal with respect to capture and survival probability is independent of the fate of any other animal.

The reduced m-array is a sufficient statistic for this model, meaning it contains all the information necessary to estimate the model parameters. Let v_i denote the number of animals released at i and never seen again, $\pi_{i,j}$ and χ_i the cell-probabilities respectively associated to the reduced m-array terms $m_{i,j}$ and v_i, and K the number of capture occasions. It follows that $\pi_{i,j} = \phi_{j-1}p_j$ for $j = i+1$, $\pi_{i,j} = \left[\prod_{t=i}^{j-2} \phi_t(1-p_{t+1})\right] \phi_{j-1}p_j$ for $j > i+1$, whilst $\chi_K = 1$ and $\chi_i = 1 - \sum_{j=i+1}^{K} \pi_{i,j}$ for $i < K$ (see for example Lebreton et al., 1992). The likelihood of the CJS model can then be expressed as:

$$L(\boldsymbol{\phi}, \mathbf{p}; \{m_{i,j}, v_i\}) = \prod_{i=1}^{K-1} \frac{\left(\sum_{j=i+1}^{K} m_{i,j} + v_i\right)!}{(m_{i,i+1})! \dots (m_{i,K})!(v_i)!} \prod_{j=i+1}^{K} \pi_{i,j}^{m_{i,j}} \chi_i^{v_i} \tag{2.1}$$

Burnham (1991) showed that the first two model assumptions are testable through a factorisation of the CJS likelihood. They are respectively the object of existing diagnostic Test 2 and Test 3, which are implemented in software such as programs RELEASE or U-CARE (Cooch and White, 2014, Chapter 5). These informative components result from the pooling-peeling algorithm adopted by Burnham (1991), that was later extended to multi-state models by Pradel et al. (2003).

2.1.3 Existing diagnostic tests

Figure 2.1 summarises the factorisation steps used by Burnham (1991), which result in two separate goodness-of-fit components and one parameter estimation component based on the minimal sufficient statistics (i.e. the minimal information needed to estimate the model parameters). The following colour coding will be used throughout this subsection for the different components:

red for parameter estimation, blue for Test 3 and green for Test 2.

The first two likelihood components, relating to first captures and losses/gains on captures, are only meaningful when simultaneously estimating abundance in an open population (see for example McCrea and Morgan, 2014, p.150). The CJS model, however, conditions on first capture and its likelihood corresponds to the probability of the recaptures given the releases. The detail of the pooling-peeling steps resulting in the final factorisation are illustrated in Figure 2.2, using an example m-array.

The notations used to express the factorised likelihood components in probabilistic terms are as follows:

- K: the number of capture occasions in the experiment,
- $\{h\}$: the previous capture history of a specific subcohort of animals,
- R_i: the number of animals released at a given occasion i,
- r_i: the number of animals released at i who are re-captured,
- $m_{i,j}$ and $\pi_{i,j}$: respectively the elements of the reduced m-array and their associated cell-probabilities as defined previously,
- $m^P_{i,j} = \sum_{k=1}^{i} m_{k,j}$: the partial column sums for $i < j$,
- $m^P_j = \sum_{i=1}^{j-1} m_{ij}$, for $j = 2, ..., K-1$: the terms peeled out during the last step of the pooling-peeling algorithm. Indeed, the first term peeled out (see Figure 2.2) is $m^P_2 = m_{1,2}$, at the next stage, the second term peeled out is $m^P_3 = m_{1,3} + m_{2,3}$ and so forth.
- $C_j = C_{j-1} - m^P_{j-1} + r_{j-1}$, for $j > 2$, with $C_2 = r_1$: the total number of animals from which the m^P_j terms are peeled out (i.e. total of the conditional multinomials pooled prior to the current peeling step). For the example used in Figure 2.2, $C_2 = r_1 = 19 + 20 + 3 + 0 + 22 = 64$, $C_3 = (64 - 19) + (19 + 11 + 35 + 0) = 110$

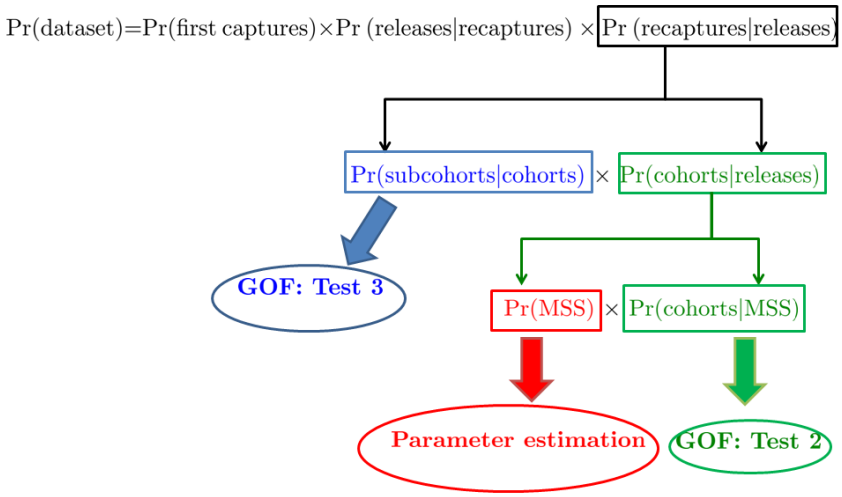


Figure 2.1: Burnham's likelihood factorisation of the CJS model, resulting in separate goodness-of-fit (GOF) and parameter estimation components, MSS denotes the minimal sufficient statistics

- $\tau_j = [\phi_{j-1} p_j] \,/\, \left[\sum_{k=j}^{K} \pi_{(j-1),k}\right]$: the conditional probabilities of the peeled bits m_j^P, for $j = 2, ..., K-1$.

The MSS for the CJS model are defined by Burnham (1991) as the r_i terms, with $i = 1, ..., K-1$ and the m_i^P terms, with $i = 2, ...K-1$. The final components of the factorised likelihood are presented below in probabilistic terms: .

$$\prod_{i=1}^{K-1} \frac{\prod_{\{h\}} \binom{R_{i;\{h\}}}{m_{i,i+1;\{h\}},...,m_{i,K;\{h\}},R_{i;\{h\}}-r_{i;\{h\}}}}{\binom{R_i}{m_{i,i+1},...,m_{i,K},R_i-r_i}} \tag{2.2}$$

$$\prod_{i=1}^{K-1} \binom{R_i}{r_i} \left(\sum_{j=i+1}^{K} \pi_{ij}\right)^{r_i} \left(1 - \sum_{j=i+1}^{K} \pi_{ij}\right)^{R_i - ri} \tag{2.3}$$

$$\prod_{j=2}^{K-1} \binom{C_j}{m_j^P} (\tau_j)^{m_j^P} (1-\tau_j)^{C_j - m_j^P} \tag{2.4}$$

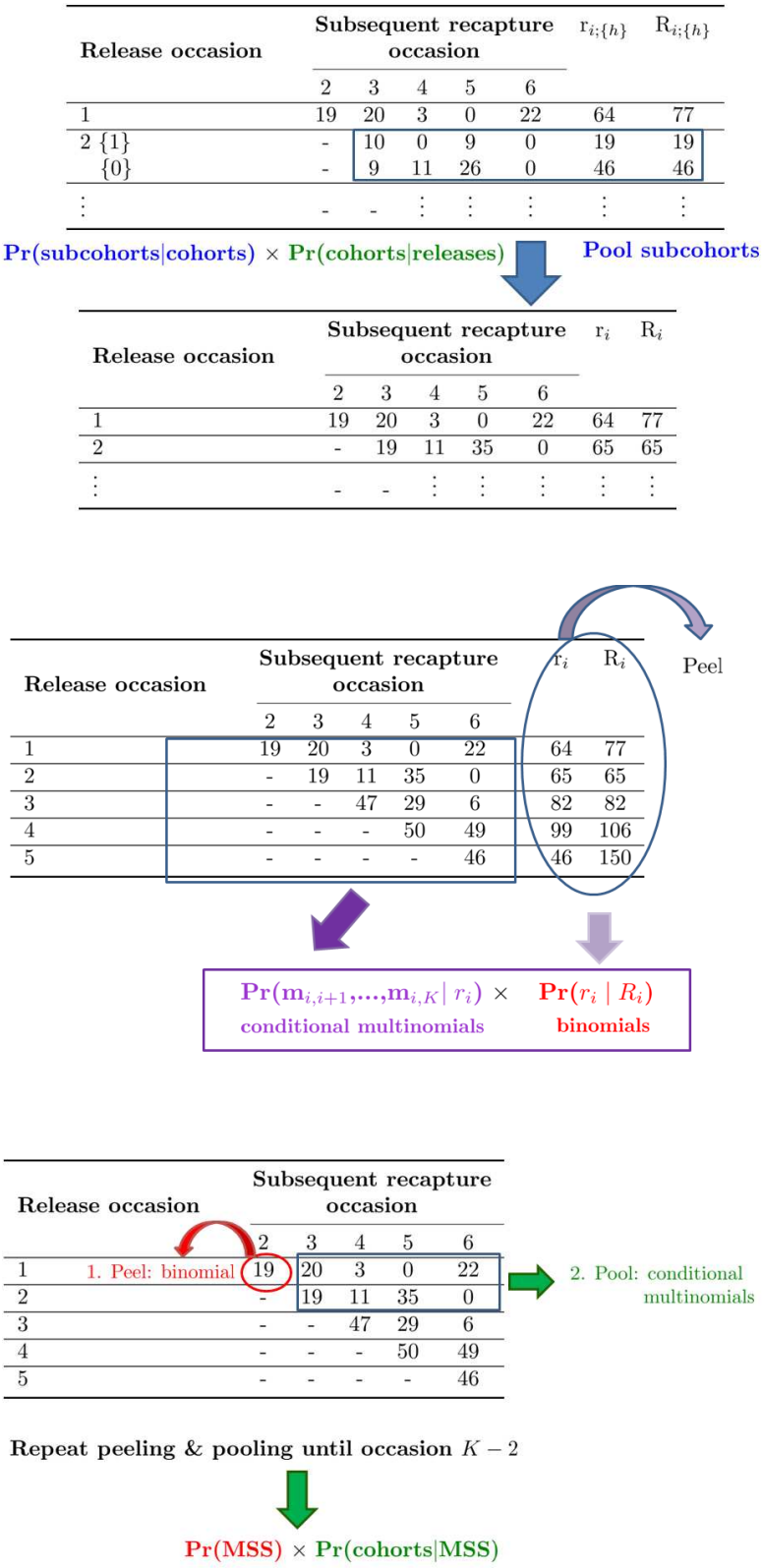


Figure 2.2: An illustration of the pooling-peeling steps leading to components Test 2 and Test 3

$$\prod_{i=2}^{K-2} \frac{\binom{C_i - m_i^P}{m_{i-1,i+1}^P, \ldots, m_{i-1,K}^P} \binom{r_i}{m_{i,i+1}, \ldots, m_{i,K}}}{\binom{C_{i+1}}{m_{i,i+1}^P, \ldots, m_{i,K}^P}} \tag{2.5}$$

Equation 2.2 represents the conditional distribution of the subcohorts given the cohorts, they consist of $K-1$ multivariate hypergeometrics and give rise to Test 3 (Burnham et al., 1987). Equation 2.3 and 2.4 denote the likelihood components used for parameter estimation: the probability of the minimal sufficient statistics. More precisely, Equation 2.3 refers to the binomials peeled off in the very first step $P(r_i \mid R_i)$ whilst Equation 2.4 refers to the binomials peeled at each occasion in the last step $P(m_i^P \mid C_i)$. Equation 2.5 represents the conditional distribution of the cohorts given the MSS, they consist of $K-3$ hypergeometrics (Pradel et al., 2003). This component is used for Test 2, to test the homogeneity between cohorts.

Test 3 and Test 2 can both be divided in subcomponents, their description and associated hypotheses are given below, based on the detailed description given in Choquet et al. (2005). Test 3 is subdivided into two components: Test 3.SR detects transience, which occurs when animals are just passing through the sampling site and hence only seen once. At each occasion, Test 3.SR compares, for animals encountered at occasion i, the newly marked animals to previously marked animals with respect to whether they are seen again. The null hypothesis for this test is H_0: "There is no difference in the probability of being seen again between newly and previously marked animals encountered at occasion i" and the alternative hypothesis is defined as $\bar{H}_0$. The specific directional departure from the null hypothesis corresponding to transience is: "Amongst the animals encountered at occasion i, the newly marked animals present a lower probability of being seen again than previously marked animals". The second component, Test 3.Sm has a less clear interpretation, it compares, for animals encountered at occasion i, the timing of recapture of the newly marked animals seen again and of the previously marked animals

that are seen again. The null hypothesis for this test is H_0: "There is no difference in the expected time of first re-encounter between newly and previously marked animals encountered at occasion i", and the alternative is defined as $\bar{H}_0$. In the same way, Test 2 is subdivided into Test 2.CT and Test 2.CL. Test 2.CT detects short-term trap-effects, which occur when capture at the current occasion affects the probability of being captured at the next occasion: trap-shyness if it decreases and trap-happiness if it increases. To identify this trap-dependence phenomenon, Test 2.CT compares animals captured at i and those not captured at i, in terms of whether they were recaptured at $i+1$ or later (given that the animals are alive at both i and $i+1$). The null hypothesis for this test is H_0: "For the animals known to be alive at both i and $i+1$, there is no difference in the probability of being seen again at $i+1$ between the animals encountered at i and those not encountered at i", and the alternative is defined as $\bar{H}_0$. The specific directional departure from the null hypothesis corresponding to trap-happiness (shyness) is: "For the animals known to be alive at both i and $i+1$, the probability of being seen again at $i+1$ is higher (lower) for the animals encountered at i than those not encountered at i". Test 2.CL, which is thought to possibly detect longer term trap-effects, compares the timing of re-encounter for the animals re-encountered after $i+1$. The null hypothesis for this test is defined as: "For the animals known to be alive at both i and $i+2$, there is no difference in the expected time of next re-encounter between the animals encountered at i and those not encountered at i", and the alternative is defined as $\bar{H}_0$.

Table 2.5: Contingency table for Test 3.SR at occasion i

	Seen Again	Never Seen Again
Newly released at i	$\sum_{j=i+1}^{K} m_{i,j;\{00\ldots0\}}$	$R_{i;\{00\ldots0\}} - \sum_{j=i+1}^{K} m_{i,j;\{00\ldots0\}}$
Previously released	$\sum_{j=i+1}^{K} \sum_{h\neq\{00\ldots0\}} m_{i,j;\{h\}}$	$\sum_{h\neq\{00\ldots0\}} R_{i;\{h\}} - \sum_{j=i+1}^{K} m_{i,j;\{h\}}$

Table 2.6: Contingency table for Test 3.Sm at occasion i

	Seen Again at $i+1$	Seen Again at ...	Seen Again at K
Newly released at i	$m_{i,i+1;\{00...0\}}$	...	$m_{i,K;\{00...0\}}$
Previously released	$\sum_{h\neq\{00...0\}} m_{i,i+1;\{h\}}$	...	$\sum_{h\neq\{00...0\}} m_{i,K;\{h\}}$

Table 2.7: Contingency table for Test 2.CT at occasion i

	Seen subsequently at $i+1$	Seen subsequently at a later occasion
Captured at i	$m_{i,i+1}$	$\sum_{j=i+2}^{K} m_{i,j}$
Not captured at i	$\sum_{l=1}^{i-1} m_{l,i+1}$	$\sum_{l=1}^{i-1}\sum_{j=i+2}^{K} m_{l,j}$

Table 2.8: Contingency table for Test 2.CL at occasion i

	Seen Again at $i+2$	Seen Again at ...	Seen Again at K
Captured at i	$m_{i,i+2}$	...	$m_{i,K}$
Not captured at i	$\sum_{l=1}^{i-1} m_{l,i+2}$	...	$\sum_{l=1}^{i-1} m_{l,K}$

These four tests are constructed from the contingency tables presented from Table 2.5 to 2.8 (Pradel et al., 2003; Choquet et al., 2009a). Firstly, standard chi-square tests of homogeneity comparing expected and observed values are derived per occasion. Then a global result may be obtained for each of the four tests by summing the independent chi-square statistics from each occasion. Finally, the sum of these four tests form the omnibus goodness-of-fit test (Choquet et al., 2005).

In practice, the data can be sparse for Test 3.Sm and/or Test 2.CL, the cells within the corresponding tables are then pooled (this does not occur for components 2.CT and 3.SR, which are based on 2×2 tables). If the data are still sparse after pooling, Fisher's exact test is used (Choquet et al., 2005).

As its name implies, Fisher's exact test computes the exact probability of obtaining the observed 2×2 table or a table more extreme, under the null hypothesis of independence (see for example Upton and Cook, 2008). The diagnostic goodness-of-fit tests are Non Applicable (NA) when the corresponding contingency tables have a row or column total of zero.

2.2 The test of positive association, a new approach for detecting heterogeneity in capture

The proposed approach stems from the observation that, if some animals have a higher capture probability than others, they will be seen more often. In such a case, at a given capture occasion i, animals with a higher number of previous encounters are also likely to present a higher number of future encounters. In other words, a positive association is expected between previous and future encounters in the presence of heterogeneity in capture.

We propose the following steps to construct the test-statistic of interest to our objective. Firstly, the test should target heterogeneity in capture and therefore should not be contaminated by noise due to deaths or permanent emigration. Hence, the occasions after the last sighting, for which the presence of the animal is uncertain, are not taken into account. Secondly, since the CJS model conditions on first capture, the period prior to the first capture and the first capture occasion itself are not informative; thus these occasions are not taken into account for the test. Similarly, the last capture occasion is not taken into account either since it does not provide any information to discriminate between the animals in terms of capture intensity. Thirdly, the occasions of first and last capture can differ amongst animals, leading to an artificial difference between them: the earlier (the later) the animals are first

(last) seen, the more possible encounters they have. Therefore, the information relative to the encounters is standardised by dividing the number of previous (future) encounters by the maximum number of possible previous (future) encounters. Fourthly, the raw proportions of previous and future encounters per animal at a given occasion are not of interest per se. Rather, we are interested in how animals fare relatively to one another: are animals that are seen more (less) often before i also seen more (less) often after i? Therefore, the ranks of these proportions constitute the final information retained from the data to test for heterogeneity in capture. Finally, since the range of ranks is limited and that we expect many ties, Goodman-Kruskal's gamma is used to test for a positive association between the ranks of previous and future encounters (Siegel and Castellan Jr., 1988, p. 291). Since the test is based on previous and future encounters with respect to a given capture occasion i, it is reasonable to require a minimum of two informative occasions (i.e. excluding the first and last sightings) both before and after i. As a result, the test is restricted to animals known to be alive at least at $i+3$ and released before $i-1$; so it can only be computed for capture-recapture experiments with at least six capture occasions. Note that the capture history information at occasion i could be counted in either the future or previous encounters; as there is no strong argument in favour of either side and we decided to count it in the previous encounters.

We use a toy example comprising three capture histories (see Table 2.9) to illustrate the construction of the test at a given occasion $i = 5$. In our example, animal ID 98 is not included within the test construction because it is released before occasion 4, but never seen again, so not known to be alive at occasion 8. The numbers of previous (future) encounters (denoted m), proportions (denoted pr) and maximum possible number of previous (future) encounters, denoted max, are shown in Table 2.9 for our example animals ID 99 and ID 100.

Table 2.9: A toy example for extracting the information required for the test of positive association: for the test per occasion, at occasion i=5 and for the global test. For the test per occasion, the occasion of interest (here i=5) is denoted in bold. For the global test, the middle occasion is denoted in bold. m denotes the number of encounters, max the maximum possible number of encounters, and pr the proportion. The information related to previous encounters and future encounters is respectively denoted in red and green.

Test of positive association per occasion

	Capture History										Previous encounters				Future encounters			
Occasion i	1	2	3	4	**5**	6	7	8	9	10	m	max	***pr***	***rank***	m	max	***pr***	***rank***
ID 98	0	1	0	0	**0**	0	0	0	0	0	Not taken into account for test (not known to be alive at occasion 8)							
ID 99	1	0	0	1	**1**	1	1	1	0	0	2	4	**2/4**	**2**	2	2	**2/2**	**2**
ID 100	0	0	1	0	**0**	0	1	0	1	0	0	2	**0/2**	**1**	1	3	**1/3**	**1**

Global test of positive association

	Capture History										Previous encounters				Future encounters			
Occasion i	1	2	3	4	5	6	7	8	9	10	m	max	***pr***	***rank***	m	max	***pr***	***rank***
ID 98	0	1	0	0	0	0	0	0	0	0	global test of positive association not applicable							
ID 99	1	0	0	**1**	1	1	1	1	0	0	1	3	**1/3**	**2**	3	3	**3/3**	**2**
ID 100	0	0	1	0	0	**0**	1	0	1	0	0	3	**0/3**	**1**	1	2	**1/2**	**1**

The gamma measure of positive association, denoted by G, is estimated by γ, which is based on the pairs of discordant D and concordant C observations:

$$\gamma = \frac{C - D}{C + D}$$

A pair of observations is concordant if the observation ranking higher (lower) for the previous encounters, also ranks higher (lower) for the future encounters; and discordant if the observation ranking higher (lower) for the previous encounters ranks lower (higher) for the future encounters. In our example from Table 2.9, animal ID 99 is ranked higher than animal ID 100 for both previous encounters and future encounters. Thus, animals ID 99 and ID 100 form a concordant pair. Animals who are ranked the same for either previous encounters or future encounters form ties and are not taken into account by the gamma measure.

In the case of heterogeneity in capture, we expect a high number of concordant pairs. Hence, we use a one-sided test with the null hypothesis defined as "G$\leq$ 0" and the alternative as "G$>$ 0". With this test, we hope to detect the specific departure of heterogeneity in capture resulting from differences in the animals' behaviour, when some present a higher capture probability than others. When the subset of animals considered for the test at occasion i, n, is relatively large, under the null hypothesis of no association, the distribution of the test-statistic $\gamma/\sqrt{Var(\gamma)}$ is approximately a standard normal (Siegel and Castellan Jr., 1988). In order to be conservative regarding this approximation, we propose to restrict n to at least 30. If $n < 30$, we state that the test is not applicable (NA).

We examined two versions of the test, based on different approaches to the variance estimation. The true variance of γ is known to be smaller than the upper bound $[n(1-\gamma^2)]\,/\,[C+D]$ (Siegel and Castellan Jr., 1988); and using this upper bound as variance estimate leads to a conservative test. We also

used an estimate of the asymptotic variance derived by Brown and Benedetti (1977), $\left[\sum_i \sum_j a_{ij}(A_{ij} - D_{ij})^2 - (4(C-D)^2/n)\right] / [(C+D)^2]$, where a_{ij} denotes the frequency cell from the contingency table of rank of proportions of previous encounters $\times$ rank of proportions of future encounters, with $A_{ij} = \sum_{k<i} \sum_{l<j} a_{kl} + \sum_{k>i} \sum_{l>j} a_{kl}$ and $D_{ij} = \sum_{k>i} \sum_{l<j} a_{kl} + \sum_{k<i} \sum_{l>j} a_{kl}$.

The subsets of animals used for the test at different occasions i are not independent, which means the results from each occasion cannot be pooled. However, if not much temporal variation is expected for the capture probability, one may consider a global test using the middle occasion between first and last capture which allows for an optimal balance between information brought by the previous and future occasions. The test procedure and restrictions are the same as the test for a given occasion i, only i will be replaced by the middle occasion and each animal used only once within the test. The global test of positive association is also illustrated for our toy example in Table 2.9.

2.3 Other tests investigated

2.3.1 Adapted versions of the diagnostic tests

Pradel et al. (2005) observed that, based on the contingency tables used for Test 3.SR and Test 2.CT, directional tests could be used to detect transience and trap-happiness or shyness since these phenomena lead to an expected direction of departure from the CJS assumption. For instance, if there is trap-happiness, one would expect the contingency table associated with Test 2.CT to be consistently unbalanced, with more animals captured at i and seen again at $i+1$ than expected. Tables 2.10 and 2.11 show the signs of expected unbalances in the associated contingency tables under, respectively, transience and trap-happiness. The directional components by occasion z_i are obtained by using the square-root of the chi-square statistics by occasion from

Test 2.CT or Test 3.SR, and signing them, for all occasions, according to the unbalance of interest : for example, using a positive sign if there are more animals captured at i and seen again at $i+1$ than expected. Recall that the contingency tables per occasion for Tests 2.CT and 3.SR are 2×2, the associated chi-square statistic has therefore one degree of freedom and the z_is follow a standard normal distribution. An overall directional component z may be derived for both tests with $z = \sum_{i=1}^{p} z_i/\sqrt{p}$ following a standard normal distribution, with p the number of directional components per occasion z_i.

Table 2.10: Directional test 3.SR, expected direction of departure under transience, occasion i

	Seen Again	Never Seen Again
Newly released at i	-	+
Previously released	+	-

Table 2.11: Directional test 2.CT, expected direction of departure under trap-happiness, occasion i

	Seen subsequently at $i+1$	Seen later
Captured at i	+	-
Not captured at i	-	+

When there is heterogeneity in detection, it is known that Tests 2.CT and 3.SR tend to generate significant results (Péron et al., 2010) but they do not provide a specific diagnostic of this phenomenon. For instance, a combination of trap-dependence and transience may also yield significant results for these two tests. Péron et al. (2010) suggested removing the squared overall directional components z_{3SR} and z_{2CT} of Test 3.SR and Test 2.CT from the overall omnibus chi-square statistic and using the remainder to test for

heterogeneity in capture. Denoting by Tot the omnibus statistic and dof_{Tot} the associated degrees of freedom, the test-statistic suggested by Péron et al. (2010) is $\text{TotC} = \text{Tot} - z^2_{3SR} - z^2_{2CT}$, which follows a chi-square distribution with $(dof_{\text{Tot}} - 2)$ degrees of freedom.

2.3.2 The Leslie test for equal catchability

The Leslie test for equal catchability tests whether the sampling of marked animals is random (Orians and Leslie, 1958). It is based on the frequency of recaptures, within each group of animals with the same first release occasion and last capture occasions. Using terminology from Carothers (1971), these groups are named blocks and denoted by b; and an animal may only belong to one block. As for the test of positive association, the information before and after these occasions (first and last included) is not taken into account for the test. The test requires that animals be potentially recaptured at least 3 times between their first and last capture occasion; it can therefore be used for capture-recapture experiments with at least 5 sampling occasions. The conception of the Leslie test rests on a different way of modelling the data, than the usual product-multinomial approach presented in Section 2.1.2. Orians and Leslie (1958) observed that there are two possible outcomes for each capture occasion, "recaptured (1)" or "not seen (0)"; it follows that the capture occasions constitute independent Bernoulli trials. Within each block b, the probability of success (here recapture) at each capture occasion i is the proportion of recaptures at i: $p_{i,b} = \sum_{k=1}^{N_b} CH_b(k,i)/N_b$, with $CH_b(k,i)$ denoting the ith element of the capture history of animal k belonging to block b and N_b the number of animals in that block. Let f and l respectively denote the first and last contribution times to the test of animals from a given block. (Here, f and l will respectively be the occasion following the first capture and the one preceding the last capture.) Consequently, the number of recaptures per

individual k, within b, denoted by $S_{k,b}$ follows a Poisson-Binomial distribution with mean $\sum_{i=f}^{l} p_{i,b}$.

Orians and Leslie (1958) use the ratio of observed variance over expected variance as their test-statistic per block:

$$L_b = \sum_{k=1}^{N_b} \frac{\left(S_{k,b} - \bar{S}_b\right)^2}{\sum_{i=f}^{l} p_{i,b}(1-p_{i,b})}, \text{ with } \bar{S}_b = \frac{\sum_{k=1}^{N_b} S_{k,b}}{N_b} \tag{2.6}$$

The distribution of L_b was not formally proven in Orians and Leslie (1958). However, Carothers (1971) established an equivalence between Leslie's test and Cochran's Q, thus showing that a corrected version of L_b has a proven asymptotic distribution. Cochran's Q is used to test whether matched samples of proportions differ (Siegel and Castellan Jr., 1988, p. 170). Applied to the Leslie-test framework, within block b, the animals constitute the "matched samples" whilst the capture occasions (between first and last) constitute the "subjects", since the aim is to assess whether the capture probability differs between animals. The animals are "matched" because they all present "responses" for each of the capture occasions (re-captured or not seen). The test-statistic for Cochran's Q is then given below:

$$Q = \frac{N_b(N_b-1)\sum_{k=1}^{N_b}(S_{k,b} - \bar{S}_b)^2}{N_b \sum_{i=f+1}^{l-1}\sum_{k=1}^{N_b} CH_b(k,i) - \sum_{i=f+1}^{l-1}\sum_{k=1}^{N_b}\left(CH_b(k,i)\right)^2} \tag{2.7}$$

Recall that $p_{i,b} = \sum_{k=1}^{N_b} CH_b(k,i)/N_b$, hence:

$$\begin{aligned} L_b &= \sum_{k=1}^{N_b} \frac{\left(S_{k,b} - \bar{S}_b\right)^2}{\sum_{i=f}^{l}\left[\frac{\sum_{k=1}^{N_b} CH_b(k,i)}{N_b}\left(1 - \frac{\sum_{k=1}^{N_b} CH_b(k,i)}{N_b}\right)\right]} \\ &= N_b \sum_{k=1}^{N_b} \frac{\left(S_{k,b} - \bar{S}_b\right)^2}{\sum_{i=f}^{l}\left[\sum_{k=1}^{N_b} CH_b(k,i) - \frac{\sum_{k=1}^{N_b}(CH_b(k,i))^2}{N_b}\right]} \end{aligned}$$

From this, it follows that, for each block,

$$Q = \frac{N_b - 1}{N_b} L_b \tag{2.8}$$

Since Q's asymptotic distribution was proven to be asymptotically $\chi^2_{N_b-1}$ (see for example Siegel and Castellan Jr., 1988, p. 170), we use this corrected version of the Leslie-test (or in other words Cochran's Q), keeping the original sample size recommendation of $N_b \geq 20$ per group (Orians and Leslie, 1958).

One drawback of the Leslie test is that it discards a lot of data; Carothers (1971) builds on Cochran's Q as defined above in a capture-recapture context, and extends it to make it more efficient by using more data. This extension is described in Section 2.3.3.

2.3.3 Carothers' extension

The extension of Carothers (1971) consists in making Leslie's test more efficient by using the data from all blocks instead of having one test per block. As for Leslie's test, the testable data will consist only of the segments of capture histories comprised between first and last occasion. The testable data is shown for an example capture history matrix, in Table 2.12. Carothers (1971) required that only occasions presenting a number of testable animals, H_i, of at least 20 or above, be considered to ensure the validity of the asymptotic distributions used. Note here the major difference from Leslie's test with respect to the data discarded. Indeed, Leslie's test discarded all blocks with less than 20 animals, whereas Carothers' extension only discards the capture occasions presenting less than 20 animals with testable data. The information from these discarded occasions is not taken into account in any of the quantities calculated for the test, as shown in Table 2.12 (greyed out in the table). Basically, those occasions are considered as non-existent. Similarly to Leslie's test, the animals from the testable data (minus sparse occasions) are

grouped into blocks. Only the grouping criterion differs slightly since animals are grouped by their first and last times of contribution to the test data f and l; this can differ from the grouping by first and last occasions due to the elimination of the sparse occasions. For instance, the last animal from block 1 has a different first capture occasion (2) than the other animals in the block (1), but it contributes to the test from occasion 3 to 4 as do all the other animals from this same block.

Since the test-statistic used by Carothers (1971) is not a straightforward product of the raw data, its construction steps are presented in detail in this section. Carothers (1971) considers the testable data after discarding the occasions as appropriate and makes use of yet another perspective on the CJS data. Under the assumption of equal recapture probability for all marked animals known to be alive, all the 1's occurring at a specific occasion i have the same probability of occurring across individuals (these probabilities may vary across the capture occasions). The total number of recaptures per occasion SO_i is fixed; thus the number of recaptures per occasion follows a hypergeometric distribution (sampling without replacement). Let p_i denote the recapture probability at occasion i, or in other words, the probability that element $CH(k,i)$ (ith element of the capture history of animal k) is equal to 1: $p_i = P[CH(k,i) = 1] = SO_i/H_i$. These quantities are calculated for our example in Table 2.12. Based on the properties of the hypergeometric distribution (see for example Lecoutre, 2006), we obtain: $\mathrm{E}[CH(k,i)] = p_i$ and $\mathrm{Var}[CH(k,i)] = p_i(1-p_i)$. Since the sampling occasions are independent, $\mathrm{Cov}[CH(k,i), CH(k,j)] = 0$ for $i \neq j$. However, since the number of recaptures per occasion is fixed, the sampling between individuals is not independent and the covariance between 2 animals k_1 and k_2 for a given capture

Table 2.12: Extracting the testable data and computing the quantities necessary to Carothers' test, from a raw individual capture history matrix: an example. i denotes the capture occasion, the testable data is denoted in bold and blue and the quantities calculated from the data are denoted in green.

Block b	Capture occasion i							$S_{k,b}$	N_b	$\bar{S}_b$	ϵ_b	σ_b
	1	2	3	4	5	6	7					
	0	1	0	0	0	0	0	-	-	-	-	-
1	1	**1**	**1**	**1**	1	0	0	**2**	**7**	**1.57**	**1.35**	**0.660**
	1	**1**	**1**	**1**	1	0	0	**2**				
	1	**0**	**0**	**1**	1	0	0	**1**				
	1	**1**	**0**	**0**	1	0	0	**0**				
	1	**1**	**1**	**1**	1	0	0	**2**				
	1	**0**	**1**	**1**	1	0	0	**2**				
	0	1	**1**	**1**	1	0	0	**2**				
2	0	0	1	**1**	**1**	1	0	**2**	**8**	**1.63**	**1.35**	**0.660**
	0	0	1	**1**	**1**	1	0	**2**				
	0	0	1	**1**	**1**	1	0	**2**				
	0	0	1	**0**	**1**	1	0	**1**				
	0	0	1	**1**	**1**	1	0	**2**				
	0	0	1	**0**	**1**	1	0	**1**				
	0	0	1	**1**	**1**	1	0	**2**				
	0	0	1	**0**	**1**	1	0	**1**				
3	0	1	**1**	1	0	0	0	**1**	**4**	**1**	**-**	**-**
	0	1	**1**	1	0	0	0	**1**				
	0	1	**1**	1	0	0	0	**1**				
	0	1	**1**	1	0	0	0	**1**				
4	1	**0**	**1**	**0**	**1**	1	0	**2**	**10**	**1.7**	**2.06**	**0.801**
	1	**0**	**1**	**1**	**0**	1	0	**2**				
	1	**1**	**1**	**0**	**1**	1	0	**2**				
	1	**1**	**0**	**1**	**0**	1	0	**1**				
	1	**1**	**1**	**1**	**1**	1	0	**3**				
	1	**1**	**0**	**0**	**1**	1	0	**1**				
	1	**1**	**0**	**0**	**1**	1	0	**1**				
	1	**1**	**0**	**1**	**0**	1	0	**1**				
	1	**1**	**1**	**0**	**1**	1	0	**2**				
	1	**1**	**1**	**1**	**0**	1	0	**2**				
5	0	0	0	1	**0**	1	0	**0**	**3**	**0.33**	**-**	**-**
	0	0	0	1	**0**	1	0	**0**				
	0	0	0	1	**1**	1	0	**1**				
H_i	0	16	**21**	**25**	**21**	0	0					
SO_i	-	-	**15**	**16**	**15**	-	-					
p_i	-	-	**0.71**	**0.64**	**0.71**	-	-					

occasion i is:

$$\begin{aligned}\mathrm{Cov}[CH(k_1,i),CH(k_2,i)] &= E[CH(k_1,i),CH(k_2,i)] - E[CH(k_1,i)]E[CH(k_2,i)] \\ &= \frac{SO_i}{H_i}\frac{SO_i-1}{H_i-1} - \frac{SO_i^2}{H_i^2} \\ &= \frac{SO_i(SO_i-1)H_i - SO_i^2(H_i-1)}{H_i^2(H_i-1)} \\ &= \frac{SO_i}{H_i}\left[-\frac{H_i-SO_i}{H_i}\right]\frac{1}{H_i-1}\end{aligned}$$

Hence,

$$\mathrm{Cov}[CH(k_1,i),CH(k_2,i)] = -\frac{p_i(1-p_i)}{H_i-1} \tag{2.9}$$

Recall from Section 2.3.2 that $S_{k,b}$ follows a Poisson-binomial distribution. Cochran (1950) and Carothers (1971) show that the asymptotic joint distribution of the $S_{k,b}$ is a multivariate normal. Within a block, all the $S_{k,b}$ follow a joint multivariate distribution with common expectation and variance; since $S_{k,b} = \sum_{i=f}^{l} CH_b(k,i)$, it follows that:

$$\mathrm{E}[S_{k,b}] = \sum_{i=f}^{l} p_i \tag{2.10}$$

and

$$\mathrm{Var}[S_{k,b}] = \sum_{i=f}^{l} p_i(1-p_i)\,. \tag{2.11}$$

Also, for animals k_1 and k_2 (who may belong to different blocks b_a and b_b), the covariance is:

$$\mathrm{Cov}[S_{k_1,b_a}, S_{k_2,b_b}] = \sum_{i=max(f_{b_a},f_{b_b})}^{min(l_{b_a},l_{b_b})} \frac{-p_i(1-p_i)}{H_i-1}\,. \tag{2.12}$$

In order to simplify notations, Carothers (1971) denoted $\mathrm{E}[S_{k,b}]$ by ϵ_b, $\mathrm{Var}[S_{k,b}]$ by σ_b^2; and $\mathrm{Cov}[S_{k_1,b_a}, S_{k_2,b_b}]$ by $\rho_{b_a,b_b}\sigma_{b_a}\sigma_{b_b}$, introducing the correlation between

blocks ρ_{b_a,b_b}. Using Walsh's theorem, Carothers (1971) deduced that

$$O_b = \frac{1}{\sigma_b^2(1-\rho_{b,b})}\sum_{k=1}^{N_b}(S_{k,b}-\bar{S}_b)^2 \sim \chi^2_{(N_b-1)} \tag{2.13}$$

with $\rho_{b,b}$ denoting the correlation within block b; the distributions of O_b being independent between blocks as well as from the mean number of recaptures per individual in block b, $\bar{S}_b$. He showed that the $\bar{S}_b$ have an asymptotic multivariate normal distribution and that $B-1$ of them are linearly independent hence,

$$O_R = (\bar{S}_b-\epsilon_b)^T\mathbf{\Sigma^{-1}}(\bar{S}_b-\epsilon_b) \sim \chi^2_{(B-1)} \tag{2.14}$$

with B denoting the number of blocks and $\mathbf{\Sigma}$ the variance-covariance matrix of the $\bar{S}_b$ for $b=1$ to $B-1$. The off-diagonal elements of Σ are

$$\mathrm{Cov}[\bar{S}_b,\bar{S}_c] = \rho_{b,c}\sigma_b\sigma_c \text{ for } b \neq c\,, \tag{2.15}$$

and the diagonal elements

$$\mathrm{Var}[\bar{S}_b] = \frac{1}{N_b^2}\left[\sum_{k=1}^{N_b} Var[S_{k,b}] + \sum_{k=1}^{N_b}\sum_{k\neq l}\mathrm{Cov}(S_{k,b},S_{l,b})\right] \quad = \frac{\sigma_b^2}{N_b} + \frac{N_b-1}{N_b}\rho_{b,b}\sigma_b^2\,. \tag{2.16}$$

Finally, the test-statistic used by Carothers (1971) for equal catchability is $O_T = \sum_{b=1}^{B} O_b + O_R$, which follows a χ^2 distribution with $\sum_{b=1}^{B}(N_b-1) + B - 1 = \left(\sum_{b=1}^{B} N_b\right) - 1$ degrees of freedom. Again, for reasons of asymptotic validity, Carothers (1971) recommends using only the blocks with $\bar{S}_b \geq 1.5$ for the computation of the test-statistic; for our example from Table 2.12, blocks 3 and 5 are ignored. The quantities ϵ_b and σ_b are computed using Equations 2.10 and 2.11, and presented in Table 2.12 for blocks 1, 2 and 4. For example,

for block 1,

$$\epsilon_1 = p_3 + p_4 = 0.71 + 0.64 = 1.35$$
$$\sigma_1 = \sqrt{p_3(1-p_3) + p_4(1-p_4)} = \sqrt{0.71(1-0.71) + 0.64(1-0.64)} \quad = 0.660$$

The correlation matrix between (off-diagonal terms) and within blocks (diagonal terms) is computed for the example from Table 2.12, based on Equations 2.11 and 2.12:

$$\mathbf{P} = \begin{array}{c} \\ 1 \\ 2 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 4 \end{array} \\ \begin{bmatrix} -0.044 & -0.023 & -0.036 \\ -0.023 & -0.044 & -0.036 \\ -0.036 & -0.036 & -0.045 \end{bmatrix} \end{array}.$$

For example, the correlation between blocks 1 and 2 is computed as

$$\rho_{1,2} = \frac{\frac{-p_4(1-p_4)}{H_4-1}}{\sigma_1\sigma_2} = \frac{\frac{-0.64(1-0.64)}{25-1}}{0.660 \times 0.660} = -0.023\,.$$

And the correlation within block 1 from

$$\rho_{1,1} = \frac{\frac{-p_3(1-p_3)}{H_3-1} + \frac{-p_4(1-p_4)}{H_4-1}}{\sigma_1\sigma_1} = -0.044\,.$$

In the same way, applying Equation 2.13, $O_1 = 8.17$, $O_2 = 4.12$ and $O_4 = 6.12$. To preserve linear independence, a block needs to be set aside. We follow the example from Carothers (1971) and set aside the smallest one: block 1. From Equation 2.16, $\mathrm{Var}(\bar{S}_2) = 0.038$, $\mathrm{Var}_(\bar{S}_4) = 0.038$ and from Equation 2.15,

$\text{Cov}(\bar{S}_2, \bar{S}_4)$=-0.019 resulting in: $\mathbf{\Sigma}^{-1} = \begin{bmatrix} 35.29 & 17.56 \\ 17.56 & 34.91 \end{bmatrix}$

Finally, from Equation 2.14, $O_R = 3.75$, and we have all the elements needed to compute the test statistic: $O_T = 8.17 + 4.12 + 6.12 + 3.75 = 22.16$, which produces a p-value of 0.57 for a χ^2 with 24 degrees of freedom ($N_1 + N_2 + N_4 - 1$). There is no significant evidence for heterogeneity in capture in our example dataset.

2.4 Simulation study

We used simulation to assess and compare the performance of the tests described in Sections 2.2 and 2.3.

2.4.1 Simulation scenarios

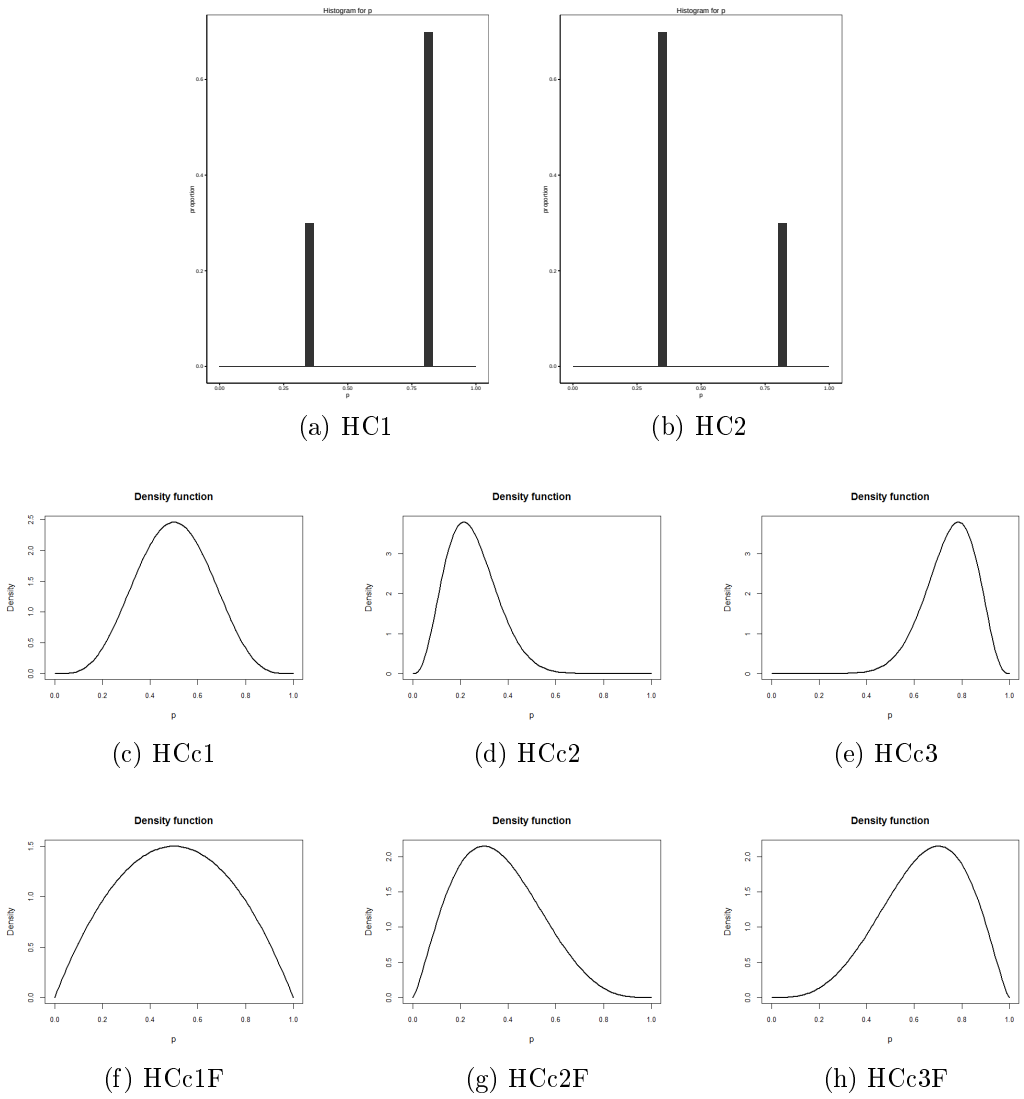
A subset of the different scenarios simulated to investigate the methods considered are shown in Table 2.13. The tests' performances were evaluated in good conditions for survival ($\phi = 0.9$). We considered control datasets (scenarios denoted by C1 and C2), generated by a CJS model with constant capture and survival probabilities, in order to check the Type I error rate obtained. Then, we assessed whether, in good conditions, the tests were powerful to detect heterogeneity in capture. Our basic heterogeneity scenarios had two groups of animals with contrasting capture probabilities of 0.35 and 0.82 and a proportion of 0.3 for either one of the groups; p_1, p_2, ϕ_1 and ϕ_2 respectively denote the capture and survival probabilities in groups 1 and 2, π_1 denotes the proportion of individuals in group 1. Based on these discrete heterogeneous capture scenarios, denoted by HC1 and HC2, we also considered a slight temporal dependence, by adding a uniform term U[-0.20,0.17] to the original capture probabilities at each time-point; we denote these scenarios by HC1t and HC2t. The discrete scenarios aimed to represent situations where

animals had intrinsically different behaviours in terms of capture (e.g. dominants and subordinates), thus movement between groups was not allowed. We also considered different cases of continuous heterogeneity in capture, with the capture probability, p, following a beta distribution:

- HCc1, symmetric around the mean, generated by a $\beta(5, 5)$: mean 0.5 and standard deviation (sd) 0.15.
- HCc2, positive skew (most animals with low capture probabilities), generated by a $\beta(4, 12)$: mean(sd) $= 0.25(0.11)$.
- HCc3, negative skew (most animals with high capture probability), generated by a $\beta(12, 4)$: mean(sd) $= 0.75(0.11)$.
- HCc1F, symmetric around the mean, generated by a $\beta(2, 2)$: mean(sd) $= 0.50(0.22)$.
- HCc2F, positive skew (most animals with low capture probabilities), generated by a $\beta(2.4, 4.3)$: mean(sd) $= 0.36(0.17)$.
- HCc3F, negative skew (most animals with high capture probability), generated by a $\beta(4.3, 2.4)$: mean(sd) $= 0.64(0.17)$.

The distributions of p for the capture heterogeneity scenarios considered are illustrated in Figure 2.3.

Finally, we wish our test to be specifically sensitive to heterogeneity in capture (represented by the scenarios described previously). Therefore, to assess the tests' specificity to heterogeneity in capture, other situations causing a violation of the CJS model assumptions were considered: short-term trap-dependence (respectively denoted by TH and TS for trap-happiness and trap-shyness), transience (denoted by TR) and heterogeneity in survival (denoted by HS), as well a combination of trap-dependence and transience (denoted by



(a) HC1 (b) HC2

(c) HCc1 (d) HCc2 (e) HCc3

(f) HCc1F (g) HCc2F (h) HCc3F

Figure 2.3: Distribution of p for heterogeneity scenarios: histograms and density plots

TSTR for trap-shyness and THTR for trap-happiness). We hope the test of positive association will be insensitive to these departures.

The scenarios of transience and short-term trap-dependence considered are described in Table 2.13: ϕ_{a1} denotes survival of newly marked animals, ϕ_{a2} the survival of previously marked animals; p_{TA} and p_{NTA} respectively denote the probability of capture of a trap-aware and non-trap-aware animal. Recall that an animal is trap-aware at a given occasion i if it has been captured at $i-1$.

For each scenario, 250 datasets were simulated for a capture-recapture experiment with 10 capture occasions, under two different sample size conditions: N = 2000 and N = 500.

2.4.2 Results

The simulation results presented are the percentage of significant test results (out of the number of cases where the test was applicable), using a 5% level.

Test of positive association

The results of the tests per occasion are presented for both the conservative version and the version based on Brown and Benedetti (1977)'s estimate of the variance, using two informative occasions before and after the occasion tested for, in Tables 2.14 to 2.17. We also explored the effect of using only one informative occasion before and after the occasion tested for, and those results are displayed in Tables 2.18 and 2.19. The results of the global test per dataset, are shown in Table 2.20 for all the test versions mentioned above.

Table 2.14 shows that, for N = 2000, the conservative test per occasion using two informative occasions presents a very small Type I error rate (lower than 5% as expected, but also close to 0). This test has very high power at all occasions for situations with discrete heterogeneity in capture (around 100% of significant results per occasion). It is also powerful for scenarios

Table 2.13: Parameter values for the simulation scenarios considered. p_1, p_2, ϕ_1 and ϕ_2, respectively denote the capture and survival probabilities in groups 1 and 2, π_1 denotes the proportion of individuals in group 1. ϕ_{a1} denotes survival of newly marked animals, ϕ_{a2} the survival of previously marked animals. p_{TA} and p_{NTA} denote the probability of capture of a trap-aware and non-trap-aware animal

Scenario	p_1	p_2	ϕ_1	ϕ_2	π_1	ϕ_{a1}	ϕ_{a2}	p_{TA}	p_{NTA}
Control:									
C1	0.35	0.35	0.9	0.9	-	-	-	-	-
C2	0.82	0.82	0.9	0.9	-	-	-	-	-
Heterogeneous capture (2 groups):									
HC1	0.35	0.82	0.9	0.9	0.3	-	-	-	-
HC2	0.35	0.82	0.9	0.9	0.7	-	-	-	-
Heterogeneous survival (2 groups):									
HS	0.9	0.9	0.45	0.9	0.3	-	-	-	-
Trap-shyness:									
TS	-	-	0.9	0.9	-	-	-	0.62	0.82
Trap-happiness:									
TH	-	-	0.9	0.9	-	-	-	0.55	0.35
Transience:									
TR	0.82	0.82	-	-	-	0.4	0.9	-	-
Trap-Shyness & Transience									
TSTR	-	-	-	-	-	0.4	0.9	0.62	0.82
Trap-Happiness & Transience									
THTR	-	-	-	-	-	0.4	0.9	0.55	0.35

of continuous heterogeneity, for scenarios HCc1F, HCc2F and HCc3F (i.e. when heterogeneity is strong) with around 70 to 100% of significant results (apart from occasion 3, for HCc2F: 47.20%). It does not perform as well for weaker scenarios of continuous heterogeneity (12 to 76.4% of significant results for scenarios HCc1, HCc2 and HCc3), although it does retain good power at several occasions for HCc1 and HCc3. For a smaller sample size (see Table 2.15), this test presents adequate power for scenarios with discrete heterogeneity (approximately 50 to 90% of significant results per occasion), but it does not perform well for continuous scenarios (mostly 10-20 % of significant results per occasion, except for HCc1F: around 60% and HCc2: consistently under 5%).

The test is only slightly sensitive to short-term trap-happiness for a couple of occasions: for example, there are 8% of significant results at occasion 7. Considering the conservative nature of this test, it is much higher than the type I error but we note that it is close to the 5% level that would be used in practice. Importantly, the test is not affected by transience or heterogeneity in survival, nor is it affected by trap-shyness. It is slightly affected by a combination of transience and trap-happiness. This is expected considering that trap-happiness on its own affects the test.

To sum up, the conservative test is powerful at detecting situations of strong heterogeneity in capture and shows promising results to specifically detect this phenomenon.

Tables 2.16 and 2.17 show that the version of this test based on the Brown & Benedetti variance estimate presents a Type I error close to the 5% level for N = 2000. It is slightly higher for N = 500 (around 6-7% for all occasions), this could be due to the distribution being slightly less well approximated than for the larger sample size of N = 2000. For N = 2000, the test based on the Brown & Benedetti variance estimate is very powerful at detecting heterogeneity in capture for the discrete cases (around 100% for all occasions) as well as most

Table 2.14: Test of positive association, conservative version, using two informative occasions, per occasion, N = 2000 animals, percentage of significant results (number of applicable tests), high percentage of significant results in bold (> 50%)

Capture occasion	3	4	5	6	7
C1	0.40 (250)	0.00 (250)	0.00 (250)	0.00 (250)	0.40 (250)
C2	0.80 (250)	0.00 (250)	0.00 (250)	0.00 (250)	0.40 (250)
HC1	**86.40** (250)	**98.80** (250)	**100.00** (250)	**100.00** (250)	**98.40** (250)
HC2	**96.00** (250)	**99.60** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)
HC1t	**92.80** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)
HC2t	**96.40** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)
HCc1	29.60 (250)	**54.40** (250)	**76.40** (250)	**73.20** (250)	**62.80** (250)
HCc2	11.60 (250)	11.20 (250)	17.20 (250)	14.40 (250)	12.00 (250)
HCc3	14.80 (250)	33.60 (250)	**53.20** (250)	**57.20** (250)	45.60 (250)
HCc1F	**89.60** (250)	**99.60** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)
HCc2F	47.20 (250)	**75.60** (250)	**90.40** (250)	**85.60** (250)	**77.20** (250)
HCc3F	**70.00** (250)	**88.00** (250)	**98.40** (250)	**98.80** (250)	**98.00** (250)
HS	1.20 (250)	0.40 (250)	0.40 (250)	0.40 (250)	0.40 (250)
TS	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)
TH	3.20 (250)	1.60 (250)	4.00 (250)	5.20 (250)	8.00 (250)
TR	0.80 (250)	0.80 (250)	0.40 (250)	0.00 (250)	0.40 (250)
TSTR	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)
THTR	1.21 (248)	0.80 (250)	0.40 (250)	5.60 (250)	4.40 (250)

Table 2.15: Test of positive association, conservative version, using two informative occasions, per occasion, N = 500 animals, percentage of significant results (number of applicable tests), high percentage of significant results in bold (> 50%)

Capture occasion	3	4	5	6	7
C1	0.00 (57)	0.40 (250)	0.40 (250)	0.00 (250)	1.21 (248)
C2	1.20 (167)	1.20 (250)	0.80 (250)	0.40 (250)	0.80 (250)
HC1	36.03 (136)	47.60 (250)	**65.60** (250)	**58.80** (250)	44.00 (250)
HC2	37.21 (86)	**62.00** (250)	**82.00** (250)	**90.40** (250)	**88.80** (250)
HC1t	**64.74** (156)	**86.40** (250)	**79.20** (250)	**83.60** (250)	**71.60** (250)
HC2t	48.72 (39)	**60.32** (247)	**80.40** (250)	**79.60** (250)	**56.05** (248)
HCc1	6.86 (102)	8.40 (250)	12.80 (250)	12.40 (250)	12.80 (250)
HCc2	0.00 (7)	4.61 (217)	3.21 (249)	2.40 (250)	3.93 (178)
HCc3	9.38 (160)	6.40 (250)	10.00 (250)	11.60 (250)	9.20 (250)
HCc1F	31.94 (72)	43.60 (250)	**64.00** (250)	**66.00** (250)	**60.40** (250)
HCc2F	12.50 (24)	13.31 (248)	20.00 (250)	19.60 (250)	21.20 (250)
HCc3F	17.83 (129)	25.60 (250)	38.00 (250)	34.00 (250)	33.20 (250)
HS	0.00 (25)	2.82 (248)	1.20 (250)	0.40 (250)	1.20 (250)
TS	0.62 (161)	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)
TH	1.52 (66)	0.40 (250)	0.80 (250)	3.20 (250)	2.40 (250)
TR	NA (0)	0.00 (71)	0.85 (236)	0.40 (249)	1.20 (250)
TSTR	NA (0)	0.00 (72)	0.00 (231)	0.00 (248)	0.00 (242)
THTR	NA (0)	0.00 (15)	0.88 (114)	1.28 (156)	2.99 (67)

of the continuous scenarios (around 100 % for strong heterogeneity scenarios HCc1F, HCc2F, HCc3F and generally more than 50% for HCc1, HCc2 and HCc3). It retains very good power for all the scenarios with strong heterogeneity for N = 500 (more than 50% for all scenarios at all occasions). Its power is lower for the continuous scenarios with weaker heterogeneity (approximately 20 to 50% for HCc1, HCc2 and HCc3).

This version of the test is also sensitive to trap-happiness (20 to 60% of significant results per occasion for N = 2000 and around 20% for N = 500), whilst it is not affected by trap-shyness. The percentage of significant results for both transience and heterogeneity in survival is close to the Type I error rate. Finally the test does not react to a combination of trap-shyness and transience but is moderately sensitive to a transience and trap-happiness combined (around 15 to 20% of significant results for both N = 2000 and N = 500).

In conclusion, the Brown & Benedetti version of the test is more powerful than the conservative test for detecting situations with weaker heterogeneity in capture. However, it also reacts strongly to trap-happiness, making it difficult to distinguish between both phenomena.

We note that the use of two informative occasions for past and future encounters results in a non-negligible loss of data for the tests of positive association: see the low number of applicable tests in the presence of transience in Table 2.15 for instance; and Figure 2.4, which presents boxplots of the number of animals actually used for the tests for simulated datasets under HC1, for both N = 2000 and N = 500. However, this restriction seems optimal for detecting heterogeneity in capture specifically, particularly with the conservative test. Indeed, if only one informative occasion is used, Tables 2.18 and 2.19 shows that the tests are much more sensitive to short-term trap-happiness (approximately 30-40% significant results per occasion for the conservative test, for N = 2000 versus a maximum of 8% on one occasion for the test with two informative occasions; 60-80% for Brown & Benedetti with one informative

Table 2.16: Test of positive association based on Brown & Benedetti's variance estimate, using two informative occasions, per occasion, N = 2000 animals, percentage of significant results (number of applicable tests), high percentage of significant results in bold (> 50%)

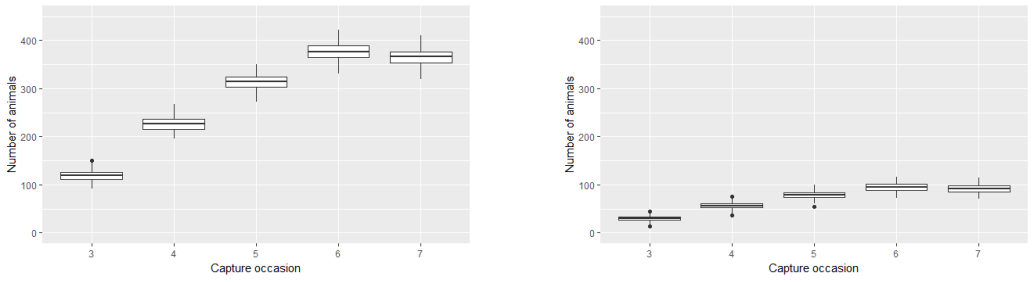
Capture occasion	3	4	5	6	7
C1	4.40 (250)	3.60 (250)	6.80 (250)	6.40 (250)	4.80 (250)
C2	4.00 (250)	6.00 (250)	2.80 (250)	4.40 (250)	3.20 (250)
HC1	**98.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)
HC2	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)
HC1t	**98.80** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)
HC2t	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)
HCc1	**76.40** (250)	**90.80** (250)	**98.40** (250)	**98.80** (250)	**94.40** (250)
HCc2	37.20 (250)	**56.00** (250)	**59.60** (250)	**58.80** (250)	44.80 (250)
HCc3	**51.20** (250)	**76.40** (250)	**86.80** (250)	**89.20** (250)	**82.80** (250)
HCc1F	**99.20** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)
HCc2F	**82.80** (250)	**98.40** (250)	**100.00** (250)	**98.80** (250)	**97.60** (250)
HCc3F	**94.00** (250)	**98.80** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)
HS	6.40 (250)	1.60 (250)	5.20 (250)	6.80 (250)	4.80 (250)
TS	1.20 (250)	0.40 (250)	0.00 (250)	0.00 (250)	0.00 (250)
TH	19.20 (250)	30.40 (250)	34.00 (250)	45.20 (250)	44.00 (250)
TR	7.60 (250)	5.60 (250)	2.40 (250)	5.20 (250)	4.80 (250)
TSTR	2.00 (250)	2.80 (250)	0.40 (250)	0.80 (250)	0.80 (250)
THTR	11.69 (248)	18.00 (250)	24.00 (250)	23.20 (250)	28.00 (250)

Table 2.17: Test of positive association based on Brown & Benedetti's variance estimate, using two informative occasions, per occasion, N = 500 animals, percentage of significant results (number of applicable tests), high percentage of significant results in bold (> 50%)

Scenario	3	4	5	6	7
C1	5.26 (57)	4.40 (250)	6.80 (250)	6.00 (250)	6.05 (248)
C2	6.59 (167)	6.80 (250)	5.60 (250)	6.40 (250)	7.60 (250)
HC1	**60.29** (136)	**84.00** (250)	**92.00** (250)	**90.00** (250)	**79.60** (250)
HC2	**81.40** (86)	**95.20** (250)	**98.00** (250)	**99.60** (250)	**97.60** (250)
HC1t	**84.62** (156)	**97.60** (250)	**92.80** (250)	**94.80** (250)	**90.40** (250)
HC2t	**84.62** (39)	**93.52** (247)	**99.20** (250)	**97.60** (250)	**89.92** (248)
HCc1	28.43 (102)	49.60 (250)	**54.40** (250)	**53.60** (250)	44.80 (250)
HCc2	0.00 (7)	22.58 (217)	24.90 (249)	20.40 (250)	17.98 (178)
HCc3	23.75 (160)	30.80 (250)	38.00 (250)	39.20 (250)	32.40 (250)
HCc1F	**75.00** (72)	**90.00** (250)	**93.20** (250)	**94.40** (250)	**88.80** (250)
HCc2F	**62.50** (24)	**54.44** (248)	**66.40** (250)	**66.40** (250)	**58.80** (250)
HCc3F	**50.39** (129)	**70.80** (250)	**79.60** (250)	**78.40** (250)	**70.80** (250)
HS	0.00 (25)	4.03 (248)	2.80 (250)	3.60 (250)	3.20 (250)
TS	2.48 (161)	0.80 (250)	3.20 (250)	0.80 (250)	0.80 (250)
TH	13.64 (66)	13.60 (250)	16.80 (250)	18.80 (250)	19.60 (250)
TR	NA (0)	9.86 (71)	5.51 (236)	6.02 (249)	4.00 (250)
TSTR	NA (0)	4.17 (72)	1.30 (231)	1.21 (248)	2.48 (242)
THTR	NA (0)	0.00 (15)	14.91 (114)	12.18 (156)	17.91 (67)

occasion versus 20-60% for two informative occasions). There is no beneficial effect of considering one informative occasion, for the tests per occasion, in regards to identifying heterogeneity in capture specifically, thus we did not deem it necessary to present the tables relevant to N = 500.

Table 2.20 presents the results obtained with the global positive association test, for all the versions of the test described above, for both N = 2000 and N = 500. As expected the global test is more powerful than the test per occasion and we find the same trends as for the tests per occasion. For the test with two informative occasions, high power is observed for both the conservative test and the Brown & Benedetti version to detect heterogeneity in capture, except for the continuous scenarios of heterogeneity when N = 500, for the conservative test. The tests are not sensitive to heterogeneity in survival nor transience or trap-shyness but are affected by trap-happiness (moderately for conservative 18%, strongly for Brown & Benedetti with 65.2% for N = 2000). The results from the global test display another argument in favour of using two informative occasions for the positive association test. Indeed when using one informative occasion, the Type I error rate for Brown & Benedetti is higher than expected, even for N = 2000 (11.60% for C2), whereas it is around 5% for C1 and C2 when using two informative occasions, indicating that the distributional properties of the test-statistic are better in this case.



(a) HC1, N = 2000, number of animals tested

(b) HC1, N = 500, number of animals tested

Figure 2.4: Test of positive association by occasion, using 2 informative occasions: boxplots of the number of animals used per test, scenario HC1

Table 2.18: Percentage of significant results (number of applicable tests), test of positive association per occasion, using one informative occasion, conservative version, N = 2000, high percentage of significant results in bold (> 50%)

Capture occasion	2	3	4	5	6	7	8
C1	1.60 (250)	0.00 (250)	0.00 (250)	0.80 (250)	0.00 (250)	0.40 (250)	0.40 (250)
C2	0.80 (250)	0.40 (250)	0.40 (250)	0.40 (250)	0.00 (250)	0.40 (250)	0.40 (250)
HC1	**82.00** (250)	**98.80** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**96.00** (250)
HC2	**76.00** (250)	**98.80** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)
HC1t	**94.80** (250)	**98.80** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**99.60** (250)
HC2t	**80.80** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)
HCc1	18.40 (250)	41.20 (250)	**59.60** (250)	**69.20** (250)	**74.40** (250)	**72.40** (250)	42.80 (250)
HCc2	7.60 (250)	16.00 (250)	14.40 (250)	20.00 (250)	18.40 (250)	14.40 (250)	8.40 (250)
HCc3	8.80 (250)	26.80 (250)	43.20 (250)	**50.40** (250)	**58.00** (250)	**52.80** (250)	37.20 (250)
HCc1F	**73.20** (250)	**97.20** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**97.60** (250)
HCc2F	32.80 (250)	**66.40** (250)	**85.20** (250)	**88.40** (250)	**88.80** (250)	**86.80** (250)	**60.80** (250)
HCc3F	47.60 (250)	**81.20** (250)	**92.80** (250)	**98.40** (250)	**98.40** (250)	**98.40** (250)	**89.20** (250)
HS	0.40 (250)	0.00 (250)	0.40 (250)	0.00 (250)	0.00 (250)	0.80 (250)	0.80 (250)
TS	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)
TH	5.60 (250)	16.00 (250)	16.40 (250)	27.60 (250)	32.80 (250)	46.80 (250)	42.40 (250)
TR	0.40 (250)	0.00 (250)	0.00 (250)	0.40 (250)	0.40 (250)	0.40 (250)	0.00 (250)
TSTR	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)
THTR	6.40 (250)	3.60 (250)	5.20 (250)	9.20 (250)	12.00 (250)	16.80 (250)	18.80 (250)

Table 2.19: Percentage of significant results (number of applicable tests), test of positive association per occasion using Brown and Benedetti's asymptotic variance and 1 informative occasion, N = 2000, high percentage of significant results in bold (> 50%)

Capture occasion	2	3	4	5	6	7	8
C1	5.20 (250)	4.40 (250)	4.80 (250)	5.60 (250)	4.00 (250)	3.60 (250)	4.40 (250)
C2	1.60 (250)	4.00 (250)	4.00 (250)	4.00 (250)	6.40 (250)	2.80 (250)	4.00 (250)
HC1	**95.60** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**99.60** (250)
HC2	**98.80** (250)	**99.60** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)
HC1t	**98.80** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**99.60** (250)
HC2t	**97.60** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)
HCc1	**51.20** (250)	**83.20** (250)	**94.00** (250)	**96.40** (250)	**97.60** (250)	**96.80** (250)	**80.80** (250)
HCc2	31.20 (250)	**50.00** (250)	**54.80** (250)	**60.00** (250)	**60.80** (250)	**53.60** (250)	30.80 (250)
HCc3	34.40 (250)	**64.00** (250)	**80.40** (250)	**85.20** (250)	**86.00** (250)	**80.80** (250)	**72.40** (250)
HCc1F	**93.20** (250)	**99.60** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)
HCc2F	**68.40** (250)	**92.40** (250)	**98.40** (250)	**99.60** (250)	**99.60** (250)	**98.80** (250)	**88.40** (250)
HCc3F	**82.00** (250)	**97.60** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**98.80** (250)
HS	2.40 (250)	5.60 (250)	1.20 (250)	5.60 (250)	4.00 (250)	4.40 (250)	3.20 (250)
TS	0.00 (250)	0.40 (250)	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)
TH	34.80 (250)	**56.80** (250)	**65.20** (250)	**78.00** (250)	**81.60** (250)	**85.20** (250)	**80.80** (250)
TR	6.80 (250)	3.60 (250)	4.80 (250)	3.60 (250)	6.00 (250)	2.40 (250)	3.60 (250)
TSTR	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)
THTR	24.40 (250)	38.00 (250)	41.20 (250)	41.60 (250)	45.60 (250)	**52.00** (250)	**55.20** (250)

Table 2.20: Global test of positive association, per occasion and global, N = 2000 animals, for the conservative test and the Brown & Benedetti version, with one or two informative occasions (denoted by IO), percentage of significant results (number of applicable tests), high percentage of significant results in bold (> 50%)

Scenario	Conservative				Brown&Benedetti			
	N = 2000		N = 500		N = 2000		N = 500	
	1 IO	2 IO	1 IO	2 IO	1 IO	2 IO	1 IO	2 IO
C1	1.60 (250)	0.00 (250)	0.40 (250)	0.40 (250)	6.80 (250)	4.80 (250)	7.20 (250)	6.40 (250)
C2	2.00 (250)	0.00 (250)	0.00 (250)	0.80 (250)	11.60(250)	4.80 (250)	10.00 (250)	6.00 (250)
HC1	**100.00** (250)	**100.00** (250)	36.80 (250)	**86.40** (250)	**100.00** (250)	**100.00** (250)	**98.00** (250)	**98.00** (250)
HC2	**100.00** (250)	**100.00** (250)	**74.00** (250)	**98.40** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)	**100.00** (250)
HC1t	**100.00** (250)	**100.00** (250)	**59.60** (250)	**91.60** (250)	**100.00** (250)	**100.00** (250)	**98.00** (250)	**99.20** (250)
HC2t	**100.00** (250)	**100.00** (250)	39.60 (250)	**88.80** (250)	**100.00** (250)	**100.00** (250)	**99.20** (250)	**100.00** (250)
HCc1	**92.00** (250)	**92.00** (250)	8.00 (250)	20.80 (250)	**99.60** (250)	**100.00** (250)	**56.80** (250)	**66.80** (250)
HCc2	39.60 (250)	31.20 (250)	3.20 (250)	3.20 (250)	**76.00** (250)	**75.20** (250)	30.40(250)	29.60 (250)
HCc3	**80.40** (250)	**72.40** (250)	6.00 (250)	12.80 (250)	**96.40** (250)	**91.60** (250)	**52.00** (250)	43.20 (250)
HCc1F	**100.00** (250)	**100.00** (250)	46.00 (250)	**84.80** (250)	**100.00** (250)	**100.00** (250)	**98.40** (250)	**98.80** (250)
HCc2F	**98.00** (250)	**99.20** (250)	16.00 (250)	34.00 (250)	**100.00** (250)	**100.00** (250)	**74.00** (250)	**77.20** (250)
HCc3F	**100.00** (250)	**100.00** (250)	22.40 (250)	**53.60** (250)	**100.00** (250)	**100.00** (250)	**88.80** (250)	**91.20** (250)
HS	1.20 (250)	0.80 (250)	1.20 (250)	2.40 (250)	13.20(250)	5.60 (250)	6.00 (250)	3.20 (250)
TS	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)	0.00(250)	0.00 (250)	0.00(250)	1.60 (250)
TH	**94.40** (250)	18.00 (250)	7.60 (250)	2.40(260)	**99.20**(250)	**65.20** (250)	**67.60** (250)	23.60 (250)
TR	0.40 (250)	0.40 (250)	1.20 (250)	0.40 (250)	10.00(250)	3.60 (250)	8.40(250)	6.40 (250)
TSTR	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)	0.00(250)	0.80 (250)	0.00(250)	1.20 (250)
THTR	**58.40** (250)	7.60 (250)	4.53 (250)	2.42 (248)	**86.80**(250)	37.60 (250)	40.40(250)	19.35 (248)

Diagnostic goodness-of-fit components

The results obtained using the existing diagnostic GOF components (3.SR, 2.CT, 3.Sm, 2.CL) and total chi-square (denoted Total), as well as the corrected tests described in Section 2.3.1 (denoted by 3.SRC, 2.CTC and TotalC) are presented in Tables 2.21 and 2.22. The Type I error rate is close to the chosen 5% level for the control situations considered, both for N = 2000 and N = 500. Tests 3.SR and 2.CT are very powerful to detect the phenomena they were designed for. Indeed, the scenarios of short-term trap-dependence and transience display around 100% of significant results for Tests 2.CT and 3.SR respectively for both N = 2000 and N = 500. Note that Test 3.SR also reacts very strongly to heterogeneity in survival (100 % and 86.8 % of significant results for N = 2000 and N = 500, respectively). Furthermore, none of the other diagnostic test-components are affected for these scenarios. On the other hand, heterogeneity in capture seems to impact all of the GOF components: for example, scenario HC1 displays more than 50 % significant results for all four of diagnostic tests, for N = 2000 (see Table 2.21). A combination of trap-dependence and transience strongly impacts both Tests 3.SR and 2.CT (100 % of significant results for both TSTR and THTR, for N = 2000). Also, amongst all the scenarios considered, Tests 3.Sm and 2.CL are impacted only for heterogeneity in capture (e.g. respectively 67.20 % and 84% of significant results for HC1 for N = 2000). However they have only little power for datasets with 500 animals (see Table 2.22), and Test 2.CL has low power for scenario HC2, even with 2000 animals (only 18.8% of significant results).

Finally, based on the simulation scenarios considered, the corrected approach suggested by Péron et al. (2010) does not result in clear-cut conclusions regarding heterogeneity in capture: for N = 2000, the corrected total chi-square is highly significant for heterogeneity in capture as well as scenarios including transience (TR and TSTR). The corrected test-components taken

individually also react similarly for different phenomena. Test 3.SRC presents 15.6% of significant results for HC2, 27.20 % for HS; and Test 2.CTC 22% for HC2 and 26.80% for TS.

Because the examined tests did not have a particularly high power for detecting even discrete heterogeneity in capture, scenarios of continuous heterogeneity were not considered here.

To sum up, Tests 3.SR and 2.CT both tend to generate significant results when there is heterogeneity in capture, but they also tend to generate significant results for combinations of trap-dependence and transience. Tests 2.CL and 3.Sm are specifically affected by heterogeneity in capture, but lack power to detect this violation.

Leslie's test of equal catchability

We present the results of the modified Leslie test described in Section 2.3.2 pooled by first release occasion (if there is at least a non-missing test result for one of those groups, otherwise the pooled test is considered NA). The results obtained with the modified version of the Leslie test are shown in Tables 2.23 and Table 2.24. The Type I error is slightly lower than 5% for the control datasets. The test is very powerful for detecting heterogeneity in capture for N = 2000, but it is also very sensitive to trap-happiness. Also, it is impractical to use for smaller datasets, since the number of applicable tests is most often null or low (see Table 2.24). This test is not sensitive to trap-shyness, heterogeneity in survival or transience.

Carothers' test

The results of the Carothers test are presented in Tables 2.25 and 2.26. In common with Leslie's test it is powerful at detecting heterogeneity in capture, whether discrete or continuous, trap-happiness and the combination of trap-happiness and transience (around 90 to 100% significant results for all the

Table 2.21: Existing GOF components and corrected tests, N = 2000 animals, percentage of significant results (number of applicable tests), high percentage of significant results in bold (> 50%)

Scenario	3.SR	2.CT	2.CL	3.Sm	Total	3.SRC	2.CTC	TotalC
C1	5.60 (250)	5.20 (250)	6.80 (250)	4.40 (250)	5.20 (250)	6.00 (250)	4.40 (250)	5.20 (250)
C2	6.80 (250)	4.40 (250)	2.00 (250)	4.40 (250)	4.00 (250)	6.00 (250)	4.00 (250)	3.20 (250)
HC1	**58.80** (250)	**100.00** (250)	**84.00** (250)	**67.20** (250)	**100.00** (250)	10.40 (250)	21.20 (250)	**95.20** (250)
HC2	**76.80** (250)	**100.00** (250)	18.80 (250)	**68.00** (250)	**100.00** (250)	15.60 (250)	22.00 (250)	**71.20** (250)
HS	**100.00** (250)	4.40 (250)	0.80 (250)	4.80 (250)	**100.00** (250)	27.20 (250)	4.00 (250)	11.60 (250)
TS	6.00 (250)	**100.00** (250)	5.20 (250)	4.40 (250)	**100.00** (250)	6.00 (250)	26.80 (250)	12.00 (250)
TH	5.20 (250)	**100.00** (250)	6.40 (250)	6.40 (250)	**100.00** (250)	6.40 (250)	30.80 (250)	14.40 (250)
TR	**100.00** (250)	2.40 (250)	0.00 (250)	4.00 (250)	**100.00** (250)	**93.60** (250)	3.20 (250)	**56.00** (250)
TSTR	**100.00** (250)	**100.00** (250)	2.00 (250)	5.20 (250)	**100.00** (250)	**96.00** (250)	11.20 (250)	**62.00** (250)
THTR	**100.00** (250)	**100.00** (250)	6.00 (250)	4.40 (250)	**100.00** (250)	28.00 (250)	16.40 (250)	21.60 (250)

Table 2.22: Percentage of significant results (number of applicable tests), existing GOF components and corrected tests, N = 500 animals, high percentage of significant results in bold (> 50%)

Scenario	3.SR	2.CT	2.CL	3.Sm	Total	3.SRC	2.CTC	TotalC
C1	4.00 (250)	5.60 (250)	0.00 (250)	5.20 (250)	4.00 (250)	5.20 (250)	6.00 (250)	3.60 (250)
C2	6.00 (250)	5.20 (250)	3.60 (250)	6.80 (250)	4.80 (250)	6.40 (250)	5.20 (250)	4.40 (250)
HC1	24.00 (250)	**90.00** (250)	29.60 (250)	16.40 (250)	**87.20** (250)	8.80 (250)	11.60 (250)	34.40 (250)
HC2	27.20 (250)	**82.80** (250)	8.40 (250)	21.60 (250)	**78.40** (250)	10.00 (250)	7.20 (250)	18.80 (250)
HS	**86.80** (250)	1.20 (250)	0.00 (225)	5.20 (250)	**51.20** (250)	10.80 (250)	0.40 (250)	2.40 (250)
TS	5.20 (250)	**100.00** (250)	0.40 (250)	6.80 (250)	**98.00** (250)	4.40 (250)	4.80 (250)	2.40 (250)
TH	4.00 (250)	**99.60** (250)	5.60 (250)	6.00 (250)	**93.60** (250)	3.60 (250)	10.00 (250)	8.00 (250)
TR	**100.00** (250)	0.80 (250)	0.82 (244)	4.40 (250)	**100.00** (250)	20.80 (250)	1.60 (250)	3.20 (250)
TSTR	**100.00** (250)	**87.20** (250)	0.40 (250)	6.80 (250)	**100.00** (250)	24.40 (250)	3.20 (250)	7.60 (250)
THTR	**100.00** (250)	**84.00** (250)	4.40 (250)	4.80 (250)	**100.00** (250)	3.60 (250)	4.80 (250)	4.40 (250)

Table 2.23: Modified Leslie's test, N = 2000 animals, percentage of significant results (number of applicable tests), high percentage of significant results in bold (> 50%)

1st release occasion	1	2	3	4	5	6
C1	3.21 (249)	2.40 (250)	4.00 (250)	2.40 (250)	2.00 (250)	1.60 (250)
C2	1.60 (250)	4.80 (250)	2.40 (250)	2.00 (250)	2.40 (250)	1.20 (250)
HC1	**99.20** (250)	**99.20** (250)	**99.60** (250)	**97.20** (250)	**90.80** (250)	**70.80** (250)
HC2	**100.00** (250)	**99.60** (250)	**100.00** (250)	**99.60** (250)	**99.60** (250)	**86.00** (250)
HC1t	**99.60** (250)	**100.00** (250)	**98.00** (250)	**97.60** (250)	**95.20** (250)	**74.40** (250)
HC2t	**100.00** (250)	**100.00** (250)	**100.00** (250)	**99.60** (250)	**96.80** (250)	**80.40** (250)
HCc1	**85.60** (250)	**82.80** (250)	**74.40** (250)	**60.80** (250)	42.40 (250)	22.80 (250)
HCc2	36.06 (208)	31.12 (241)	28.11 (249)	22.80 (250)	17.60 (250)	8.00 (250)
HCc3	**70.80** (250)	**63.60** (250)	**58.00** (250)	47.20 (250)	27.60 (250)	18.00 (250)
HCc1F	**98.80** (250)	**99.20** (250)	**99.20** (250)	**97.60** (250)	**91.20** (250)	**58.80** (250)
HCc2F	**87.45** (247)	**88.00** (250)	**82.00** (250)	**70.00** (250)	48.40 (250)	23.60 (250)
HCc3F	**96.80** (250)	**98.00** (250)	**96.80** (250)	**88.00** (250)	**75.60** (250)	44.80 (250)
HS	5.20 (250)	4.00 (250)	4.00 (250)	1.20 (250)	1.20 (250)	1.60 (250)
TS	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)	0.00 (250)
TH	**62.40** (250)	**74.40** (250)	**79.20** (250)	**74.40** (250)	**67.60** (250)	38.80 (250)
TR	2.02 (248)	2.80 (250)	2.80 (250)	3.60 (250)	2.00 (250)	2.80 (250)
TSTR	0.00 (227)	0.00 (240)	0.00 (249)	0.00 (250)	0.00 (250)	0.00 (250)
THTR	35.94 (64)	40.78 (103)	34.18 (158)	35.52 (183)	24.12 (228)	25.51 (247)

corresponding scenarios for N = 2000). It is not sensitive to trap-shyness nor transience alone, or heterogeneity in survival. Unlike Leslie's test, it retains a very high power for a smaller sample size (see Table 2.26). Finally, the Type I error rate was around 5% or less depending on the control scenario considered, for N = 2000, but it was 10.40% for N = 500 (scenario C1). Again, this might be indicative of the asymptotic distributions involved in the test might not be as well approximated for N = 500.

2.5 Applications

We applied the tests examined in this chapter to two real-life datasets: Great cormorants and Sandwich terns. For demonstrative purposes, we present the results of both the test of positive association per occasion and the global test. However, since these tests are not independent, a choice should be made in

Table 2.24: Percentage of significant results (number of applicable tests), modified Leslie's test, N = 500 animals, high percentage of significant results in bold (> 50%)

1st release occasion	1	2	3	4	5	6
C1	0.00 (1)	NA (0)	NA (0)	0.00 (1)	0.00 (2)	0.00 (14)
C2	2.63 (76)	3.25 (123)	5.68 (176)	3.47 (202)	3.83 (235)	2.87 (244)
HC1	**82.14** (28)	**66.67** (42)	**67.62** (105)	**56.12** (139)	42.86 (189)	19.53 (215)
HC2	NA (0)	NA (0)	NA (0)	**100.00** (1)	**100.00** (1)	**100.00** (2)
HC1t	**92.22** (90)	**79.53** (127)	**81.77** (181)	**69.01** (213)	**57.85** (242)	35.12 (242)
HC2t	NA (0)	NA (0)	NA (0)	NA (0)	**100.00** (3)	0.00 (2)
HCc1	NA (0)	NA (0)	**100.00** (1)	NA (0)	0.00 (2)	0.00 (2)
HCc2	NA (0)	NA (0)	NA (0)	NA (0)	NA (0)	NA (0)
HCc3	28.57 (49)	17.33 (75)	15.89 (151)	18.67 (166)	11.26 (222)	6.96 (230)
HCc1F	**83.33** (6)	**100.00** (4)	**76.92** (13)	**63.64** (33)	49.02 (51)	24.05 (79)
HCc2F	NA (0)	NA (0)	0.00 (1)	0.00 (1)	40.00 (5)	8.33 (12)
HCc3F	**72.73** (11)	37.04 (27)	**53.85** (65)	32.76 (116)	23.64 (165)	14.29 (203)
HS	0.00 (7)	20.00 (15)	0.00 (22)	0.00 (46)	1.59 (63)	2.50 (80)
TS	0.00 (15)	0.00 (31)	0.00 (39)	0.00 (82)	0.00 (107)	0.00 (120)
TH	NA (0)	NA (0)	0.00 (4)	46.15 (13)	28.57 (28)	16.67 (48)
TR	NA (0)	NA (0)	0.00 (1)	0.00 (1)	0.00 (6)	12.50 (8)
TSTR	NA (0)	NA (0)	NA (0)	NA (0)	NA (0)	0.00 (5)
THTR	NA (0)	NA (0)	NA (0)	NA (0)	NA (0)	NA (0)

Table 2.25: Carothers' test, N = 2000 animals, percentage of significant results (number of applicable tests), high percentage of significant results in bold (> 50%)

Scenario	%(N)
C1	5.20 (250)
C2	1.20 (250)
HC1	**100.00** (250)
HC2	**100.00** (250)
HC1t	**99.20** (250)
HC2t	**100.00** (250)
HCc1	**100.00** (250)
HCc2	**87.60** (250)
HCc3	**95.20** (250)
HCc1F	**100.00** (250)
HCc2F	**100.00** (250)
HCc3F	**100.00** (250)
HS	3.31 (242)
TS	0.00 (250)
TH	**100.00** (250)
TR	1.20 (250)
TSTR	0.40 (250)
THTR	**99.20** (250)

Table 2.26: Percentage of significant results (number of applicable tests), Carothers' test, N = 500 animals, high percentage of significant results in bold (> 50%)

Scenario	% (N)
C1	10.40 (250)
C2	2.40 (250)
HC1	**95.20** (250)
HC2	**100.00** (250)
HC1t	**96.40** (250)
HC2t	**99.60** (250)
HCc1	**82.40** (250)
HCc2	**63.20** (250)
HCc3	46.80 (250)
HCc1	**99.60** (250)
HCc2	**94.40** (250)
HCc3	**96.80** (250)
HS	4.00 (250)
TS	1.60 (250)
TH	**85.60** (250)
TR	2.80 (250)
TSTR	4.40 (250)
THTR	**58.40** (250)

practice and the global test run if only little temporal variation is expected or if the data is too sparse to run the test per occasion. We chose to present only the results for the conservative version of the positive association test, since the Brown & Benedetti version essentially presents the same properties as the Carothers test.

2.5.1 Great cormorants

The data on Great cormorants were collected by the National Environmental Research Institute at Aarhus University, Denmark. We use the data on breeders only, from the period 1981-1993, collected from six different colonies (McCrea and Morgan, 2014, p. 3). Resighting effort is known to be highest at the largest colony: Vörso (VO), where the resighting conditions were the best (Hénaux et al., 2007). Therefore, we use the data pooled on all colonies as an artificial example to check whether the tests detect heterogeneity in capture,

in real-life conditions. We then focus on the cormorants first released and only ever resighted in VO during the period of interest, in order to investigate the presence of heterogeneity in capture within this colony, using the different approaches.

The results obtained using the different approaches, with a 5% level, are presented in Table 2.27 for all colonies pooled and for birds first captured and only ever resighted in VO. For all colonies pooled, the test of positive association yields a significant result at all occasions, the global positive association test result is also significant. Both approaches agree and indicate heterogeneity in capture. The diagnostic GOF tests indicate that the dataset presents transience and trap-dependence. Also, the Test 3.Sm result is significant while the Test 2.CL result is at the limit of significance. This suggests possible heterogeneity in capture. Leslie's test is NA in most cases due to sample size issues. Carothers' test result is significant, suggesting that the dataset presents trap-happiness or heterogeneity in capture.

For VO only, the test of positive association yields a significant result at occasion 7 and a result close to significance at occasion 6. Since the test is conservative, this suggests some heterogeneity in capture within VO. The significant result obtained for the global test also supports this conclusion. The results of the GOF tests are similar to those for the pooled colonies, except that the component 2.CL is no longer close to significance. Again, Leslie's test is NA in most cases and Carothers' test indicative of trap-happiness or heterogeneity in capture.

While the detection of heterogeneity within the pooled data is probably largely explained by the difference in detectability on the different colonies and by the fact that the birds tend to use the colonies unequally, the conservative test of positive association indicates that heterogeneity in capture is also present within VO. Hence, models accounting for heterogeneity in cap-

ture should be considered at the colony level. Such models have not been previously considered for these data.

2.5.2 Sandwich terns

The data on Sandwich terns were collected by the National Environmental Research Institute at Aarhus University together with the Copenhagen Bird Ringing Centre at the Danish Zoological Museum. The study took place on Hirsholm, a 15 ha inhabited island in northern Kattegat, Denmark (7 km NE of Frederikshavn; 57°29′N-10°37′E). The study of Sandwich terns is based on summarised yearly resighting data, over the period 2003-2012, of individuals ringed with small metal-rings engraved with unique numbers. The reading of the codes on this type of ring requires optimal conditions, i.e. proximity and good light. The ring-readings were not made inside the breeding colony of the Sandwich terns because the birds were nesting at a very high density inside a large colony of black-headed gulls *(Chroicocephalus ridibundus)*. Instead the codes on the rings were read when the birds were roosting or preening in the immediate proximity of the colony. The major disadvantage of carrying out the resightings on these birds was that not all of these individuals were actively engaged in a breeding attempt in the local colony. Thus some of the individuals were non-breeding birds that visited the colony, e.g. as prospectors, others were individuals that had stopped over before moving on to settle as breeders in another colony and others were individuals that turned up as visitors after having failed their breeding attempt in another colony. One of the goals of the study was to estimate survival in order to explore whether survival increased after the introduction of a control programme of large gulls, which predated on breeding adult Sandwich terns, their eggs and their chicks. Due to the large array of possible behaviours of the terns on which ring-readings were made, heterogeneity in capture was considered extremely likely; it was important

Table 2.27: Great cormorants test results (NA for Leslie's test if number of animals per group lower than 20, NA for positive association test if number of animals at given occasion lower than 30), d.o.f. denotes degrees of freedom, n denotes the number of animals used for the positive association test, significant results in bold.

Test		All colonies pooled			VO only		
Positive association	**capture occasion**	**test-statistic**	**n**	**p-value**	**test-statistic**	**n**	**p-value**
	3	NA	8	NA	NA	6	NA
	4	1.81	43	**0.035**	0.75	30	0.226
	5	2.52	73	**0.006**	0.86	52	0.194
	6	2.41	97	**0.008**	1.24	80	0.108
	7	2.88	83	**0.002**	2.02	69	**0.022**
	8	2.74	67	**0.003**	0.89	48	0.186
Global positive association	-	3.67	176	**<0.001**	1.88	130	**0.030**
Diagnostic GOF	**component**	**test-statistic**	**d.o.f.**	**p-value**	**test-statistic**	**d.o.f.**	**p-value**
	3.SR	99.84	9	**<0.001**	51.31	9	**<0.001**
	3.Sm	51.55	19	**<0.001**	27.41	15	**0.026**
	2.CT	42.33	8	**<0.001**	31.00	8	**<0.001**
	2.CL	23.65	15	0.071	11.24	9	0.259
	Total	217.36	51	**<0.001**	120.97	41	**<0.001**
	3.SR corrected	24.43	8	**0.002**	14.09	8	0.079
	2.CT corrected	11.35	7	0.120	8.77	7	0.270
	Total corrected	110.97	49	**<0.001**	61.51	39	**0.012**
Leslie's test	**1st capture occasion**	**test-statistic**	**d.o.f.**	**p-value**	**test-statistic**	**d.o.f.**	**p-value**
	1	NA	NA	NA	NA	NA	NA
	2	NA	NA	NA	NA	NA	NA
	3	NA	NA	NA	NA	NA	NA
	4	70.08	40	**0.002**	40.41	19	**0.003**
	5	NA	NA	NA	NA	NA	NA
	6	NA	NA	NA	NA	NA	NA
	7	32.72	21	**0.049**	32.72	21	**0.049**
Carothers' test		**test-statistic**	**d.o.f.**	**p-value**	**test-statistic**	**d.o.f.**	**p-value**
	-	519.56	262	**<0.001**	328.51	190	**<0.001**

to reveal its presence so as to avoid drawing erroneous conclusions regarding survival.

The results obtained using the different approaches, with a 5% level, are presented in Table 2.28 for the Sandwich tern dataset. The test of positive association yields a significant result at all occasions except for year 7, the global positive association test result is also significant. This is indicative of heterogeneity in capture. The diagnostic GOF tests indicate that the dataset presents transience and trap- dependence. Also, the Test 3.Sm result is significant while the Test 2.CL result is at the limit of significance. This, again, is indicative of heterogeneity in capture. Leslie's test is NA in most cases due to sample size issues. Carothers' test result is significant, suggesting the presence of trap-happiness or heterogeneity in capture.

The results obtained confirm the initial expectation of heterogeneous recapture within the Sandwich terns. In practice, different selection criteria were applied to the data in order to focus on more homogeneous groups of birds, for instance by minimising the risk of including individuals that were not engaged in a breeding attempt in the study colony in the specific year of study.

2.6 Discussion

In this chapter, we proposed a test of positive association based on Goodman-Kruskal's gamma as a new method for specifically detecting heterogeneity in capture, within a CJS framework. We examined different versions of the test: per occasion and global, using an upper bound variance estimate or the asymptotic variance estimate derived by Brown & Benedetti. We also investigated the effects of heterogeneity in capture on the current routinely used diagnostic goodness-of-fit tests associated with the CJS model: Tests 2.CT, 2.CL, 3.SR and 3.Sm. Finally, we also considered the Leslie and Carothers' tests of equal catchability.

Table 2.28: Sandwich terns test results (NA for Leslie's test if number of animals per group lower than 20, NA for positive association test if number of animals at given occasion lower than 30), d.o.f. denotes degrees of freedom, n denotes the number of animals used for the positive association test, significant results in bold.

Test		Sandwich terns results		
Positive association	**capture occasion**	**test-statistic**	**n**	**p-value**
	3	1.98	97	**0.024**
	4	2.10	115	**0.018**
	5	1.92	121	**0.027**
	6	1.87	119	**0.031**
	7	0.70	89	0.243
Global positive association	-	0.36	182	**0.002**
Diagnostic GOF	**component**	**test-statistic**	**d.o.f.**	**p-value**
	3.SR	127.27	8	**<0.001**
	3.Sm	38.13	17	**0.002**
	2.CT	140.02	7	**<0.001**
	2.CL	21.66	14	0.086
	Total	327.08	46	**<0.001**
	3.SR corrected	8.10	7	0.32
	2.CT corrected	8.58	6	0.20
	Total corrected	76.47	44	**0.002**
Leslie's test	**1st capture occasion**	**test-statistic**	**d.o.f.**	**p-value**
	1	200.92	96	**<0.001**
	2	NA	NA	NA
	3	NA	NA	NA
	4	NA	NA	NA
	5	NA	NA	NA
	6	66.71	35	**<0.001**
Carothers' test		**test-statistic**	**d.o.f.**	**p-value**
	-	519.56	262	**<0.001**

The simulation results show that none of the tests considered reacted to transience alone or heterogeneity in survival (apart from component 3.SR). As expected, the Leslie test was impractical because it discarded too much data. The interpretation of the existing diagnostic GOF tests was not straightforward when attempting to diagnose heterogeneity in capture. Indeed, according to the simulated scenarios examined, the simultaneous significance of Tests 2.CT, 2.CL, 3.SR and 3.Sm may indicate heterogeneity in capture. However, simulation has shown that there is relatively good power only when the sample size is very large and even then, Test 2.CL can lack power for certain situations of strong discrete heterogeneity.

Carothers' test and the Brown & Benedetti version of the test of positive association were both very powerful at detecting heterogeneity in capture, even for weak heterogeneity scenarios. However, they did not allow to distinguish between immediate trap-happiness and heterogeneity in capture. The conservative version of the test of positive association, on the other hand, performed well for scenarios of strong heterogeneity but not for scenarios of weaker heterogeneity. However, it reacted much more strongly to heterogeneity in capture than short-term trap-happiness. Thus, it seemed to us at this stage, to be the optimal compromise for diagnosing heterogeneity in capture specifically. The test of positive association is advantageous in that, unlike other approaches based on model comparison (Cubaynes et al., 2012), there is no need to make any assumptions about the model nor about the form of heterogeneity considered. It is also easier to understand from a theoretical point of view than, say, the Carothers test.

Based on the results obtained, there seems to be a trade-off between power to detect heterogeneity in capture and sensitivity to trap-happiness. Trap-happiness increases the chances of concordant pairs whilst trap-shyness increases the chances of discordant pairs, especially for short sequences of previous and future encounters. Thus trap-happiness increases the chance of pos-

itive association and trap-shyness diminishes it. As a result, trap-happiness may be confounded with heterogeneity, whilst the test is not sensitive to trap-shyness.

If not much temporal variation is expected in capture, it is advisable to use the global test of positive association since it provides a single result while being more powerful than the tests per occasion. The global test behaved well under a couple of specific scenarios of discrete heterogeneity involving additive time dependence. However, more investigation is warranted to determine how the global test might be affected by temporal variation, in particular whether strong time-dependence could affect the test in the same way as heterogeneity in capture. Therefore if strong temporal variation in capture is expected throughout the experiment (and that the sample size allows it), the test per occasion should be used.

If the test of positive association yields a significant result (even at just one occasion in the case of the test per occasion), models accounting for heterogeneity in capture should be considered at the model-building and model selection stage. Some possible techniques to incorporate heterogeneity in capture are: using observed covariates for modelling the capture probability, using a latent structure: finite mixture models, (Pledger et al., 2003) or hierarchical classes of animals with proportional capture probabilities (Oliver et al., 2011); Corkrey et al. (2012) provide a method to incorporate heterogeneity in capture in a Bayesian framework.

Also, the causes of heterogeneity in capture should be investigated from a biological perspective. This may lead to the identification of individuals with different behavioural patterns or indicate whether an adjustment to sampling is necessary. For example, a high degree of heterogeneity in capture may indicate that mixtures of breeders and non breeders are being sampled. If the group of interest is the breeders, the sampling process might be adjusted (e.g. sub-site or years selected to maximize the representation of breeders),

or the data might be cleaned post-hoc by applying strict criteria. Another possibility would be to create a smaller pilot study and adjust the sampling process accordingly on the final large study.

The discrete simulation scenarios have focused on clear situations of heterogeneity with no movements between groups. However, in real life, heterogeneity scenarios can be more complex, sampling may interact with behaviour in complex ways which blur the line between trap-dependence and heterogeneity in capture. For instance, when the cause of heterogeneous capture is the location, such as for the black-headed gulls (Prévot-Julliard et al., 1998), the birds may move between groups with low or high resighting propensity, and this would be statistically indistinguishable from trap-happiness if the location information is unavailable. If the locations are known, then spatial capture-recapture models can be considered.

Regarding the positive association test itself, note that we standardised the information regarding intensity of capture by using proportions of previous and future encounters. It might be of interest to explore a weighting system that would take into account the amount information brought by the animal. This would especially allow to distinguish between the extreme proportions: for instance a proportion of 1/1 from a proportion of 5/5, which is more informative. In this chapter we have focused on the significance of tests, we could also investigate the relationship between degree of heterogeneity and gamma estimate values.

In addition to this, the number of animals used for the test per occasion is relatively low compared to the original sample size, we therefore propose to derive an empirical p-value from a non-parametric permutation test when the data is too sparse to use the normal approximation. For the analysed dataset, the ranks of previous encounter proportions are fixed, and the gamma test-statistic computed on 10 000 distinct permutations of the ranks of future encounter proportions, in order to compute the empirical probability of obtaining

a result as extreme or more extreme than the one observed on the analysed dataset. Note that if it is not possible to obtain 10 000 distinct permutations due to low numbers, the maximum number of distinct permutations is used and we then derive the exact probability of obtaining a result as extreme or more extreme than the observed one. Based on 250 simulated datasets of smaller sample size (150 animals) under control scenario C2, using a 5% level, the permutation test per occasion results in a Type I error rate close to the expected 5%, with, respectively for occasions 3 to 7: 11.69%, 4.40%, 6.07%, 1.62% and 6.43%. For 250 simulated datasets of the same sample size, based on the heterogeneity in capture scenario HC2, the percentage of significant results obtained for occasions 3 to 7 is, respectively: 37.5%, 58.00%, 71.49%, 74.66% and 59%. Hence, based on the situations considered, the permutation test shows promising results for smaller sample sizes, behaving as expected under the control scenario C2 and showing adequate power to detect heterogeneity in capture under scenario HC2; and it should be explored further.

The test of positive association was explored for open populations in a CJS framework. But it could also be used in a context of population abundance estimation. Indeed the Jolly-Seber (JS) model, used to estimate abundance, assumes that unmarked and marked animals behave the same (McCrea and Morgan, 2014, p.149). Applied to a JS context, if the test of positive association for marked animals yields a significant result, then the assumption of equal catchability is violated. Therefore the population as a whole may exhibit heterogeneous behaviour.

Finally, the test of positive association and the Carothers test can both be used for closed population models, which do not allow for births or deaths. In this case, the animals are known to be alive during the whole experiment. Therefore, the whole encounter history becomes informative, including the information prior to the first capture occasion and after the last capture occasion.

A limitation of the test of positive association is that in a multi-state context, it could not be easily extended to detect heterogeneity in capture when the capture probabilities are state-dependent. However, it extends naturally to the study of transition probabilities between states, and this will be addressed in Chapter 3.

Chapter 3

Detecting a mover-stayer structure

3.1 Introduction

The CJS model described in Chapter 2 can be too restrictive in its assumptions, and limited in its biological scope because it utilises only the information of whether the animal is captured or not (Lebreton et al., 2009). But ecologists can be interested in aspects such as the geographic location for migratory birds, or their breeding status (see for example Hénaux et al., 2007). Such information is easily collected upon the animals' capture and recorded as states.

States are defined by Lebreton and Pradel (2002) as 'any mutually exclusive and identifiable events in the life cycle of the population under study'. They can be static, in which case they remain the same throughout the individual history: for example, sex (although occasionally it may change over time, for certain species of fish or plants for example). States can also be dynamic, in which case they either follow a deterministic (e.g. age) or stochastic (e.g. health status) process. For static and deterministic states, information is available as long as the animal is captured at least once (Lebreton et al.,

2009). The general definition of states allows multi-state models to tackle a large array of biological questions, they also have the ability to handle the more complex stochastic dynamic states.

The information of interest is collected when the animal is captured, resulting in individual capture histories that follow the same structure as in the CJS framework. The 1s are replaced by the state in which the animal is captured. For example *A B 0 C C* codes the information: "captured at occasion 1 in state A, recaptured at 2 in B, not captured at 3, recaptured in C at 4 and 5". The data may be summarised by a multi-state m-array (see Figure 3.1), which is a natural extension of the CJS m-array (McCrea and Morgan, 2014, p.88).

The Arnason-Schwarz model is the direct multi-state extension of the CJS model from a biological perspective (Pradel et al., 2003). It is conditional on the first release of individuals and relies on the same assumptions of homogeneity of the survival and recapture probabilities, for all animals in a given state r. Thus, all animals in state r at time i are assumed to have equal probability of surviving from occasion i to $i+1$, and it is denoted by ϕ_i^r. All animals in state r at time i are also assumed to have the same recapture probability, denoted by p_i^r. Furthermore, multi-state models also involve transition or movement probabilities between states; these are denoted by ψ_i^{rs}, and represent the probability of moving to state s by occasion $i+1$ for an animal in state r at time i. The processes of survival and transition are separated under the assumption that survival only depends on the state of the animal at time i (Cooch and White, 2014, Chapter 10); in simpler terms, animals survive first and then move. Animals in a given state at a given occasion are assumed to also have homogeneous behaviour in terms of transitions; the process governing the transition between states is assumed to be first-order Markovian i.e. the future state depends only on the current state (no memory). In addition to the states defined in the experiment, "dead" constitutes a state of its own

Table 3.1: Reduced multi-state m-array example, for a capture-recapture experiment with 3 states and 4 occasions; the terms $\mathrm{m}_{i,j}^{r,s}$ denote animals released at i in state r and next recaptured at j in state s and the v_i^r terms denote the animals released at i in state r and never seen again

m_{12}^{AA}	m_{12}^{AB}	m_{12}^{AC}	m_{13}^{AA}	m_{13}^{AB}	m_{13}^{AC}	m_{14}^{AA}	m_{14}^{AB}	m_{14}^{AC}	v_1^A
m_{12}^{BA}	m_{12}^{BB}	m_{12}^{BC}	m_{13}^{BA}	m_{13}^{BB}	m_{13}^{BC}	m_{14}^{BA}	m_{14}^{BB}	m_{14}^{BC}	v_1^B
m_{12}^{CA}	m_{12}^{CB}	m_{12}^{CC}	m_{13}^{CA}	m_{13}^{CB}	m_{13}^{CC}	m_{14}^{CA}	m_{14}^{CB}	m_{14}^{CC}	v_1^C
-	-	-	m_{23}^{AA}	m_{23}^{AB}	m_{23}^{AC}	m_{24}^{AA}	m_{24}^{AB}	m_{24}^{AC}	v_2^A
-	-	-	m_{23}^{BA}	m_{23}^{BB}	m_{23}^{BC}	m_{24}^{BA}	m_{24}^{BB}	m_{24}^{BC}	v_2^B
-	-	-	m_{23}^{CA}	m_{23}^{CB}	m_{23}^{CC}	m_{24}^{CA}	m_{24}^{CB}	m_{24}^{CC}	v_2^C
-	-	-	-	-	-	m_{34}^{AA}	m_{34}^{AB}	m_{34}^{AC}	v_3^A
-	-	-	-	-	-	m_{34}^{BA}	m_{34}^{BB}	m_{34}^{BC}	v_3^B
-	-	-	-	-	-	m_{34}^{CA}	m_{34}^{CB}	m_{34}^{CC}	v_3^C

and is not observable in a capture-mark-recapture context; the CJS model can then be expressed as a multi-state model with states "alive" and "dead". For convenience, the parameters of a multi-state model are usually presented in a matrix format, each time-dependent parameter being an $(S+1)\times(S+1)$ matrix for an experiment with S "live" states: $\boldsymbol{\Phi_t}$ denotes the survival matrix, $\boldsymbol{\Psi_t}$ the transition matrix and $\boldsymbol{P_t}$ the capture matrix. All these matrices are row-stochastic, meaning that the sum of probabilities in each row sums to one. An illustration is given below ("dead" is represented by †):

$$\boldsymbol{\Phi_t} = \begin{array}{c} \\ (1) \\ (2) \\ (...) \\ (S) \\ \dagger \end{array} \begin{array}{c} \begin{array}{ccccc} (1) & (2) & (...) & (S) & \dagger \end{array} \\ \begin{bmatrix} \phi_t^1 & 0 & \dots & 0 & 1-\phi_t^1 \\ 0 & \phi_t^2 & \vdots & 0 & 1-\phi_t^2 \\ \vdots & \dots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \phi_t^S & 1-\phi_t^S \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \end{array}, \quad \boldsymbol{\Psi_t} = \begin{array}{c} \\ (1) \\ (2) \\ (...) \\ (S) \\ \dagger \end{array} \begin{array}{c} \begin{array}{ccccc} (1) & (2) & (...) & (S) & \dagger \end{array} \\ \begin{bmatrix} \psi_t^{1,1} & \psi_t^{1,2} & \dots & 1-\sum_{s=1}^{S-1}\psi_t^{1,s} & 0 \\ \psi_t^{2,1} & \psi_t^{2,2} & \dots & 1-\sum_{s=1}^{S-1}\psi_t^{2,s} & 0 \\ \vdots & \dots & \dots & \vdots & \vdots \\ \psi_t^{S,1} & \psi_t^{S,2} & \dots & 1-\sum_{s=1}^{S-1}\psi_t^{S,s} & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \end{array},$$

$$\boldsymbol{P_t} = \begin{array}{c} \\ (1) \\ (2) \\ (...) \\ (S) \\ \dagger \end{array} \begin{array}{c} \begin{array}{ccccc} (1) & (2) & (...) & (S) & 0 \end{array} \\ \begin{bmatrix} p_t^1 & 0 & \dots & 0 & 1-p_t^1 \\ 0 & p_t^2 & \vdots & 0 & 1-p_t^2 \\ \vdots & \dots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & p_t^S & 1-p_t^S \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \end{array}.$$

Due to the complex nature of multi-state models, possible departures from model assumptions are numerous. The current diagnostic goodness-of-fit suite developed by Pradel et al. (2003) consists of Test 3G and Test M. Test 3G is partitioned into components used to detect memory (Test WBWA) and transience (Test 3G.SR) (Pradel et al., 2005). Memory occurs when the state reached at $i+1$ is influenced by the state at $i-1$. Test M is used to detect trap-effects.

This chapter focusses on a particular case of heterogeneous transition behaviour, where animals exhibit a mover-stayer structure, with some animals that have a tendency to move whilst the others tend to remain where they are. The mover-stayer structure was first introduced in social studies, to describe different patterns of industrial mobility (see for example Spilerman, 1972). It was originally defined as the existence of two types of individuals: some who stay where they are and others who move homogeneously. We define it more loosely here as some individuals being more likely to move (movers) than others (stayers). We assume that this behaviour is an intrinsic characteristic of the animals and therefore does not change over time. To the best of our knowledge, this phenomenon has rarely been analysed in a capture-recapture framework; but could provide new insight into biological processes such as migration patterns or disease progression stages. The detection of a mover-stayer structure naturally lends itself to an extension of the test of positive association developed in Chapter 2, to a multi-state framework.

Amongst the existing diagnostic tests, only Test WBWA is related to tran-

sitions, and it is presented in detail alongside Test 3G, from which it derives, in Section 3.2. Test M is used to test whether animals not captured at i and known to be alive are consistent with being a mixture of the animals observed in any of the states at the same occasion and known to be alive (Pradel et al., 2003). This test will be presented in detail in Chapter 4. It is not relevant to the current chapter's objective, which pertains to information on movement and therefore requires animals to be captured. Section 3.3 describes the test of positive association derived to detect a mover-stayer structure. We assess the performance of this test in detecting and identifying a mover-stayer structure using simulation and compare it to Test WBWA in Section 3.4. We find from the simulation study that our test of positive association is sensitive to memory, whereas Test WBWA reacts to phenomena other than memory. Consequently, we adapt the tests considered to make them more specific and in doing so, devise a combination of tests that allows a mover-stayer structure to be distinguished from short-term memory. The explored adaptations and their simulation results are presented in Section 3.5. The tests examined in this chapter are then applied, in Section 3.6, to the famous Canada geese (*Branta canadensis*) dataset (Hestbeck et al., 1991), which has often been used as a case study to illustrate memory. Finally we conclude in Section 3.7.

3.2 Test WBWA, a subcomponent of test 3G

Unlike the CJS model, the likelihood of the Arnason-Schwarz model cannot be expressed as a simple product: indeed if an animal is known to survive after i but is not captured at that occasion, one has to account for all the possible states to/from which it could have moved, thus introducing more complex terms in the likelihood. This can be easily illustrated through a simple example: suppose an individual presents encounter history *A 0 A* in a multi-state format, which is equivalent to *1 0 1* in a CJS format. The associated

likelihood using a CJS model would be $\phi_1(1-p_2)\phi_2 p_3$ whereas for a multi-state framework, it becomes $\phi_1^A \left[\psi_1^{AA}(1-p_2^A)\phi_2^A\psi_2^{AA} + \psi_1^{AB}(1-p_2^B)\phi_2^B\psi_2^{BA}\right] p_3^A$.

The diagnostic tests derived for multi-state models are based on the Jolly-Movement model (see for example McCrea and Morgan, 2014, p.175), rather than the Arnason-Schwarz model. Indeed, it is a more natural extension of the CJS model from a methodological point-of-view (Pradel et al., 2003). The Jolly-Movement model is only slightly different from the Arnason-Schwarz model, with the capture probabilities depending not only on the state in which the animal is captured, but also on its previous state. Test 3 (described in Chapter 2) is extended to the multi-state framework in a very straightforward manner, giving rise to Test 3G which examines whether previously marked animals and newly marked animals encountered in a given state, at a given occasion, behave in the same way. Test 3G is based on animals encountered in a given state r at a given occasion i, this test was partitioned into different informative components by Pradel et al. (2005): Test 3G.SR is associated with transience, Test WBWA is indicative of memory, and Test 3G.Sm is formed from the tables left over after partitioning (Choquet et al., 2005). The partitioning of the contingency tables associated with Test 3G is illustrated in Figures 3.1 and 3.2. As for the CJS model, Test 3G.SR examines whether previously marked and newly marked animals are as likely to be seen again. The null hypothesis associated to this test for occasion i and state r is H_0: "For animals encountered in a given state r at a given occasion i, there is no difference in the probability of being seen again later between previously marked and newly marked animals". The alternative hypothesis is defined as $\bar{H}_0$ and the specific departure of interest for transience is: "For animals encountered in a given state r at a given occasion i, the probability of being seen again later is lower for newly marked animals than for previously marked animals" (Choquet et al., 2005). Test WBWA assesses whether animals are more likely to be next re-encountered in the same state as the one they were

last seen in. The null hypothesis associated to this test for occasion i and state r is H_0:"For animals encountered in a given state r at a given occasion i, there is no difference in the expected state of next re-encounter between the animals last seen in the different states". The alternative hypothesis is defined as $\bar{H}_0$ and the specific departure of interest for memory is "For animals encountered in a given state r at a given occasion i, the probability of being next re-encountered in the same state as the one they were last seen in is higher than the probability of being next re-encountered in other states". The tests performed on the contingency tables are the usual chi-square tests of independence and Fisher's exact test is used in cases of small numbers.

Alternatively, Pradel et al. (2005) proposed using Cohen's kappa to detect memory. Cohen's kappa is typically used as a measure of agreement between two raters classifying subjects according to the same scale, taking into account the agreement that can occur by chance; it is applied to data formatted as a square contingency table (Everitt, 1992, p.146). In the memory context, the kappa is used to measure the agreement between previous and future state for animals seen at occasion i in a given state r. It is applied to the square $S \times S$ contingency tables WBWA(i,r) presented in Figure 3.2. Let a_{bs} denote the cell frequencies of table WBWA(i,r), $\mathrm{a}_{b.}$ the row sums and $\mathrm{a}_{.s}$ the column sums, and n the total number of animals in the contingency table. The proportion of agreement observed is derived from the diagonal elements of the table: $\mathrm{P}_A = \sum_{i=1}^{S} a_{ii}/n$. The proportion of agreement that may occur by chance is computed from the row and column sums: $\mathrm{P}_C = \left[(\sum_{i=1}^{S} a_{i.} a_{.i})/n \right] n$. Finally, Cohen's Kappa is defined as $\kappa = (\mathrm{P}_A - \mathrm{P}_C)/(1 - \mathrm{P}_C)$ (Everitt, 1992, p.148).

Test 3G(i,r)

	Next seen at $i+1$ in A	Next seen at $i+1$ in B	Next seen at ... in ...	Next seen at K in S	Never seen again
Last seen in A	...	...	...	...	...
⋮	⋮	⋮	⋮	⋮	⋮
Last seen in S	...	...	...	...	...
Newly marked	...	...	...	...	...

	Next seen at $i+1$ in A	Next seen at $i+1$ in B	Next seen at ... in ...	Next seen at K in S	Never seen again
Last seen in A	...	...	...	...	...
⋮	⋮	⋮	⋮	⋮	⋮
Last seen in S	...	...	...	...	...

A

	Next seen at $i+1$ in A	Next seen at $i+1$ in B	Next seen at ... in ...	Next seen at K in S	Never seen again
Previously marked	...	...	...	...	...
Newly marked	...	...	...	...	...

B

	Next seen at $i+1$ in A	Next seen at $i+1$ in B	Next seen at ... in ...	Next seen at K in S	Never seen again
Previously marked	...	...	...	...	...
Newly marked	...	...	...	...	...

B

Test 3G.Sm.a(i,r)

	Next seen at $i+1$ in A	Next seen at $i+1$ in B	Next seen at ... in ...	Next seen at K in S
Previously marked	...	...	...	...
Newly marked	...	...	...	...

Test 3G.SR(i,r)

	Seen again	Never seen again
Previously marked	...	...
Newly Marked	...	...

Figure 3.1: An illustration of the partitioning of Test 3G for animals encountered at occasion i in state r (denoted 3G(i,r)) into the informative component 3G.SR(i,r) and left-over 3G.Sm.a(i,r) components, for a capture-recapture experiment with K sampling occasions and S observable states

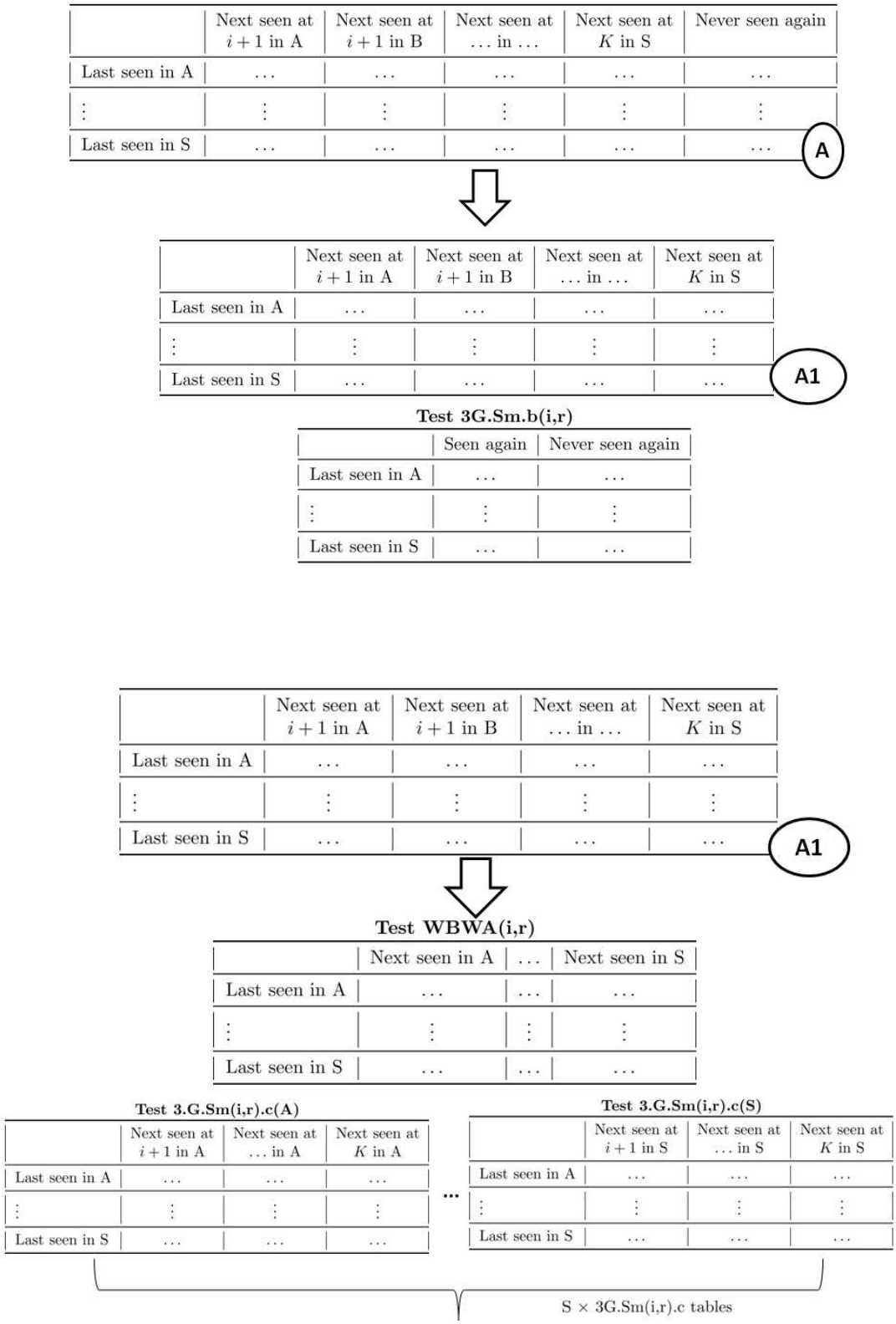


Figure 3.2: An illustration of the partitioning of Test 3G for animals encountered at occasion i in state r (denoted 3G(i,r)) into the informative component WBWA(i,r) and left-over 3G.Sm.b(i,r) and 3G.Sm.c(i,r) components, for a capture-recapture experiment with K sampling occasions and S observable states

3.3 A test to detect a mover-stayer structure within multi-state capture-recapture data

The existence of a mover-stayer structure in the population lends itself naturally to an extension of the test of positive association developed in Chapter 2. Indeed, if some individuals move more than others, we expect that, at a given occasion, animals who move more (less) before are also likely to move more (less) after. Since the question of interest is movement, occasions where the animal is not captured are not considered (they are uninformative); our test is conditional on capture. The first step is therefore to consider only the non-zero part of the capture history. We position ourselves at the optimal occasion within the non-zero capture history, in terms of information brought by previous and future movements: the middle occasion i. The capture histories are then grouped by the state in which the animal is captured at i in order to reduce noise due to state-specific properties of the animals. Also we keep only animals with at least one informative movement on each side of i; this means animals have to be captured at least three times to be part of the tested data. Then, the number of observed movements between first release occasion and the middle occasion i and the number of observed movements between i and the last capture occasion are both counted. In order to standardise this information, we use the proportion of previous and future movements, using as denominator the maximum number of previous/future possible movements conditional on capture i.e. the maximum number of previous/future movements that we could potentially observe. Finally, the ranks of these proportions are used to represent the intensity of movement of the animals relative to one another. We present a worked example of test construction in Table 3.3, based on a toy example of multi-state capture histories presented in Table 3.2. Note that animals with IDs 3 and 5 are not used for the test because they are captured less than three times.

Table 3.2: Example capture histories

Occasion	1	2	3	4	5	6	7	8	9	10
ID 1	3	0	3	3	3	0	3	0	0	0
ID 2	1	1	3	3	3	3	3	3	3	3
ID 3	3	0	0	0	0	0	0	0	0	0
ID 4	2	1	1	3	1	2	1	3	2	1
ID 5	3	2	0	0	0	0	0	0	0	0
ID 6	3	2	1	0	1	1	2	3	2	0

Table 3.3: Example capture histories: extracting the information required for the positive association test by state at the middle occasion. The middle occasion is denoted in bold for each capture history. NM denotes the number of movements, max the maximum possible number of observed movements, pr the proportion and r the rank.

	Non-zero Capture History										Previous movements				Future movements			
											NM	max	$\boldsymbol{pr}$	r	NM	max	$\boldsymbol{pr}$	r
State 1																		
ID 6	3	2	1	**1**	1	2	3	2			2	3	2/3	1	3	4	3/4	1
ID 4	2	1	1	3	**1**	2	1	3	2	1	3	4	3/4	2	5	5	1	2
State 2																		
No capture histories																		
State 3																		
ID 1	3	3	**3**	3	3						0	2	0	1	0	2	0	1
ID 2	1	1	3	3	**3**	3	3	3	3	3	1	4	1/4	2	0	5	0	1

In the same way as for Chapter 2, the range of ranks is limited and many ties are expected, so Goodman-Kruskal's gamma is used as a measure of positive association between previous and future movements. For the worked example presented in Table 3.3, the gamma estimate is not applicable for state 2 as there are no capture histories with state 2 at the middle occasion, and state 3 which presents only one tied pair. The individuals in state 1 form a concordant pair.

Similarly to Chapter 2, we expect a high number of concordant pairs for a mover-stayer structure and hence use a one-sided test with the null hypothesis defined as "G$\leq$ 0" and the alternative as "G$>$ 0". The test-statistic used is $z_s = \frac{\gamma_s}{\sqrt{\text{Var}(\gamma_s)}}$, where s denotes the state in which the animal is at the middle occasion. In the same way as for heterogeneity in capture, we investigated both the conservative version of the test and the Brown & Benedetti version. In order to be conservative regarding the distributional approximation of the test-statistic, the number of animals used for the test at each state was required to be at least 30 for the test to be applicable.

We investigated different versions of a positive association test between ranks of previous movements and ranks of future movements, split by state:

- Using animals with at least 1 informative movement on each side of the middle occasion, i.e. captured at least 3 times (see Table 3.3).
- Using animals with at least 2 informative movements on each side of the middle occasion, i.e. captured at least 5 times.

We also investigated the performance of two global versions of the test per state:

- A test over all states using the middle occasion of the capture histories, without grouping the data by state, which we expected to be sensitive to state-specific properties and thus, non optimal.

- A summary test over the states, based on the standardised sum of the independent test-statistics obtained from the test by state, with S denoting the number of states: $z_G = \sum_{s=1}^{S} \frac{z_s}{\sqrt{S}}$

3.4 Simulation study

3.4.1 Simulation scenarios

Different scenarios were explored in order to assess the performance of the different versions of the test. For each simulation scenario explored, 250 datasets of 2500 animals were simulated, with 10 capture occasions (250 animals released per occasion) and 3 live states, all equally likely to be the state at first capture. The survival probability was constant over time and states, and set to $\phi = 0.9$ for all scenarios, the same applies to the capture probability ($p = 0.9$). The simulated datasets were all generated using R.

First, we simulated simple homogeneous scenarios: only animals who tended to move (denoted M, and MO for a more extreme situation with a null probability of remaining in the same state), or only animals who tended to remain where they were (S). We then investigated whether the test could potentially react to more complex homogeneous scenarios, such as: preference (P) for one state (e.g. very high probability of moving to or remaining in State 2), avoidance (A) for one state (e.g. very high probability of moving from State 2), strongly state-dependent transition probabilities (SD1 and SD3). All these homogeneous scenarios constituted controls, used to check the Type I error rate. The transition matrices for these homogeneous scenarios are detailed in Table 3.4.

Afterwards we investigated heterogeneous scenarios with 2 groups of animals per dataset presenting different behaviours in terms of transitions. The transition matrices corresponding to these scenarios are presented in Table 3.5.

Our target was the mover-stayer structure: 2 scenarios were simulated, MS1 with a proportion $\pi_1 = 0.3$ of stayers and MS2 with a proportion $\pi_1 = 0.7$ of stayers. We also investigated how the test could be affected by animals preferring different states (P2G: one group prefers state 1, the other prefers state 2) or avoiding different states (A2G: one group avoids state 1 whilst the other avoids state 2). Finally, a scenario of heterogeneity in movement (HM) was also examined, where the animals had different movement patterns but had the same movement rate, and therefore did not present a mover-stayer structure. Apart from the mover-stayer scenarios, the heterogeneous scenarios were simulated with an equal proportion of animals from each group ($\pi_1 = 0.5$).

In addition to these heterogeneous scenarios, the more complex memory phenomenon was also examined in scenarios Mem1 and Mem2. For both Mem1 and Mem2, the probability of being at $i+1$ in the same state as at $i-1$ is higher than others, Mem2 differs from Mem1 in that, for animals in a given site s, the probability of being in the same site twice within the triplet (previous site, current site, future site) is also higher than being in each site only once. The transition matrices generating the datasets with memory under both scenarios considered are presented in Table 3.6. All these scenarios constitute violations of the multi-state model assumption of homogeneity in transitions; and their object was to assess the specificity of the test of positive association to a mover-stayer structure. Indeed, we wish our test to be sensitive to the specific departure from the null hypothesis corresponding to a mover-stayer structure, and insensitive to all other departures.

Since the transition matrix generating datasets with memory is less straightforward than the other situations, its structure is presented below, in the case of 3 live states, for transitions constant over time. Note that instead of the usual $\psi^{r,s}$, the transition probabilities are denoted by $\psi^{b,r,s}$ since the transition probability to state s at $i+1$ also depends on the state b where the animal was at $i-1$. The rows of the transition matrix are the couples of (b,r) and

the columns the couples of (r, s), whilst † represents the state "dead" (see for example Rouan et al., 2009):

$$\boldsymbol{\Psi_t} = \begin{array}{c} \\ (1,1) \\ (1,2) \\ (1,3) \\ (2,1) \\ (2,2) \\ (2,3) \\ (3,1) \\ (3,2) \\ (3,3) \\ \dagger \end{array} \begin{array}{c} \begin{array}{ccccccccccc} (1,1) & & (1,2) & (1,3) & (2,1) & (2,2) & & (2,3) & (3,1) & (3,2) & (3,3) & \dagger \end{array} \\ \begin{bmatrix} (1-\psi_t^{112}-\psi_t^{113}) & \psi_t^{112} & \psi_t^{113} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \psi_t^{121} & (1-\psi_t^{121}-\psi_t^{123}) & \psi_t^{123} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \psi^{131} & \psi_t^{132} & (1-\psi_t^{131}-\psi_t^{132}) & 0 \\ (1-\psi_t^{212}-\psi_t^{213}) & \psi_t^{212} & \psi_t^{213} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \psi_t^{221} & (1-\psi_t^{221}-\psi_t^{223}) & \psi_t^{223} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \psi_t^{231} & \psi_t^{232} & (1-\psi_t^{231}-\psi_t^{232}) & 0 \\ (1-\psi_t^{312}-\psi_t^{313}) & \psi_t^{312} & \psi_t^{313} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \psi_t^{321} & (1-\psi_t^{321}-\psi_t^{323}) & \psi_t^{323} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \psi_t^{331} & \psi_t^{332} & (1-\psi_t^{331}-\psi_t^{332}) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}.$$

Finally, for some of the scenarios, we examined the potential effect of state-dependent capture probabilities (setting $p^1 = 0.9$, $p^2 = 0.35$, $p^3 = 0.7$ with the superscripts corresponding to the states), lower capture probability ($p = 0.5$), or slightly time-dependent probabilities (adding a random uniform term $[-0.075; 0.075]$ to the probabilities of moving from a particular state). These scenarios were respectively denoted by subscripts p_s, p_L or t.

3.4.2 Main results

Test of positive association

In this section, we present the results obtained using 2 versions of the conservative positive association test split by state (simple upper bound for variance estimate): based on one informative movement on each side (i.e. animals captured at least 3 times) in Table 3.7 and based on 2 informative movements on each side in Table 3.8 (i.e. animals captured at least 5 times). We also present the results obtained using the summary test in Table 3.9, based on both versions of the test by state. All the versions of the test were coded using R. The results are presented in terms of percentage of significant test results,

Table 3.4: Transition matrices for the homogeneous simulation scenarios considered

Transition Matrices for homogeneous simulation scenarios
Movers only/Stayers only
MO $\begin{bmatrix} 0 & 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ M $\begin{bmatrix} 0.2 & 0.4 & 0.4 & 0 \\ 0.35 & 0.3 & 0.35 & 0 \\ 0.45 & 0.45 & 0.1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ S $\begin{bmatrix} 0.9 & 0.05 & 0.05 & 0 \\ 0.1 & 0.8 & 0.1 & 0 \\ 0.075 & 0.075 & 0.85 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
Preference/Avoidance for one of the states
P $\begin{bmatrix} 0.1 & 0.8 & 0.1 & 0 \\ 0.06 & 0.92 & 0.02 & 0 \\ 0.2 & 0.7 & 0.1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ A $\begin{bmatrix} 0.4 & 0.1 & 0.5 & 0 \\ 0.6 & 0.06 & 0.34 & 0 \\ 0.55 & 0.1 & 0.35 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
Strongly state-dependent transitions
SD1 $\begin{bmatrix} 0.1 & 0.45 & 0.45 & 0 \\ 0.1 & 0.8 & 0.1 & 0 \\ 0.25 & 0.25 & 0.5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ SD3 $\begin{bmatrix} 0.8 & 0.2 & 0 & 0 \\ 0.1 & 0.7 & 0.2 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

Table 3.5: Transition matrices for heterogeneous scenarios considered

Transition Matrices for heterogeneous simulation scenarios		
Mover-Stayer structure		
	Stayers	Movers
MS1, MS2:	$\begin{bmatrix} 0.9 & 0.05 & 0.05 & 0 \\ 0.1 & 0.8 & 0.1 & 0 \\ 0.075 & 0.075 & 0.85 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.2 & 0.4 & 0.4 & 0 \\ 0.35 & 0.3 & 0.35 & 0 \\ 0.45 & 0.45 & 0.1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
Preference: 2 groups		
	Prefer 1	Prefer 2
P2G:	$\begin{bmatrix} 0.91 & 0.05 & 0.04 & 0 \\ 0.7 & 0.2 & 0.1 & 0 \\ 0.84 & 0.08 & 0.08 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.1 & 0.75 & 0.15 & 0 \\ 0.04 & 0.9 & 0.06 & 0 \\ 0.06 & 0.82 & 0.12 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
Avoidance: 2 groups		
	Avoid 1	Avoid 2
A2G:	$\begin{bmatrix} 0.1 & 0.55 & 0.45 & 0 \\ 0.05 & 0.35 & 0.60 & 0 \\ 0.15 & 0.45 & 0.40 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.6 & 0.15 & 0.25 & 0 \\ 0.42 & 0.08 & 0.5 & 0 \\ 0.40 & 0.22 & 0.38 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
Heterogeneity in movement: 2 groups		
	Group 1	Group 2
HM:	$\begin{bmatrix} 0.3 & 0.4 & 0.3 & 0 \\ 0.35 & 0.5 & 0.15 & 0 \\ 0.45 & 0.45 & 0.1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.3 & 0.1 & 0.6 & 0 \\ 0.1 & 0.5 & 0.4 & 0 \\ 0.8 & 0.1 & 0.1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

Table 3.6: Transition matrices for memory scenarios considered

Transition Matrices for memory simulation scenarios
Memory

Mem1:

$$\begin{bmatrix}
\mathbf{0.7} & 0.15 & 0.15 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \mathbf{0.6} & 0.2 & 0.2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \mathbf{0.65} & 0.2 & 0.15 & 0 \\
0.22 & \mathbf{0.58} & 0.20 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.1 & \mathbf{0.8} & 0.1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.2 & \mathbf{0.62} & 0.18 & 0 \\
0.09 & 0.2 & \mathbf{0.71} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.2 & 0.15 & \mathbf{0.65} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.1 & 0.1 & \mathbf{0.8} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}$$

Mem2:

$$\begin{bmatrix}
\mathbf{0.7} & 0.15 & 0.15 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \mathbf{0.6} & 0.3 & 0.1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \mathbf{0.65} & 0.1 & 0.25 & 0 \\
0.27 & \mathbf{0.58} & 0.15 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.1 & \mathbf{0.8} & 0.1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.12 & \mathbf{0.62} & 0.26 & 0 \\
0.09 & 0.2 & \mathbf{0.71} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.08 & 0.17 & \mathbf{0.65} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.1 & 0.1 & \mathbf{0.8} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}$$

using a level of 5%.

Both the one and two informative movement tests have a very high power to detect a mover-stayer structure. Table 3.7 shows that, when one informative movement is used, 100% of the results for the tests split by state are significant for MS1 and MS2 as well as $MS1_t$ and $MS2_t$, and around 90 to 100% for the versions of these scenarios with a state-dependent or lower capture probability. However, it is slightly too sensitive in some of the control situations (e.g. 16.8% for SD1, state 1; 13.6% for P, state 1) and it does not allow us to distinguish a mover-stayer structure from short-term memory, which also results in 100% of significant results, for both Mem1 and Mem2.

When using two informative movements (see Table 3.8), the Type I error is under 5% for all control scenarios. Again, around 100% of the results for the tests split by state are significant for MS1 and MS2 as well as $MS1_t$ and $MS2_t$, the same is observed for most of the mover-stayer scenarios with a state-dependent or lower capture probability. Amongst the other heterogeneous scenarios considered, the test does not react in most cases, it is slightly sensitive only to heterogeneity in preference (6.4% for state 1) but, like the test using one informative movement, it is extremely sensitive to memory.

The results presented in Table 3.9 show that the summarised test presents the same characteristics: very powerful at detecting a mover-stayer structure and also very sensitive to memory (100% of significant results for all situations, whether one or two informative movements are used). Note that the control datasets present a Type I error lower than 5% (apart from M: 7.2% when only one informative occasion is used). Both versions of the summarised tests do not react to most of the other scenarios of heterogeneity, apart from heterogeneity in preferences scenario P2G: 17.2% for the test using two informative movements; whilst the summarised test with one informative movement is affected by time-dependence for scenario S_t: 14.0% of significant results.

Since short-term memory is a more local phenomenon than the mover-

stayer behaviour, we attempted to use animals with at least 3 informative previous and future movements (animals captured at least 7 times). However this resulted in a very high loss of data (only 150 animals used on average per dataset, out of 2500) while only marginally decreasing the sensitivity of the test to memory. The percentage of significant results obtained were respectively for states 1, 2 and 3

- 48.4 (250), 92 (250) and 92 (250) for scenario Mem1
- 24.4 (250), 44.8 (250) and 68.8 (250) for scenario Mem2

The loss in data resulting from using animals with at least two or three informative movements is not outweighed by any significant gain in terms of identifying the mover-stayer structure separately from memory. At this stage, the summarised test using animals with at least one informative previous and future movement seems to be the preferred option.

The results from the global test performed using the middle occasion without prior grouping of the animals by their state at that occasion, are not presented here due to its poor performance. Indeed it reacted strongly to control scenarios such as state-dependent transition scenario SD1 for example (54.4% using two informative movements, 80% using only one). Hence this test was not adequate for our objective.

Likewise, the results of the tests version using the Brown & Benedetti estimates are not presented in the thesis, although they were investigated. Again, these tests were sensitive to phenomena other than mover-stayer and memory: for example 56.8% of significant results for homogeneous scenario SD1, for state 1, when using the test split by state with one informative movement and 27.6% for the summarised test; 54.4% for the summarised test for P2G, using two informative movements.

Table 3.7: Test of positive association (conservative variance estimate), split by state, 1 informative movement (animal captured at least 3 times) , percentage of significant results (number of applicable tests), high percentage of significant results in bold (> 50%)

scenario	**% (N)**		
	State 1	**State 2**	**State 3**
M	2.0 (250)	1.2 (250)	5.2 (250)
S	2.0 (250)	1.2 (250)	0.8 (250)
MO	0.0 (250)	0.4 (250)	0.4 (250)
P	13.6 (250)	0.0 (250)	0.0 (250)
A	0.4 (250)	0.4 (250)	0.4 (250)
SD1	16.8 (250)	0.0 (250)	0.4 (250)
SD3	3.2 (250)	0.0 (250)	0.0 (250)
Mem1	**100.0** (250)	**100.0** (250)	**100.0** (250)
Mem2	**100.0** (250)	**100.0** (250)	**100.0** (250)
MS1	**100.0** (250)	**100.0** (250)	**100.0** (250)
MS2	**100.0** (250)	**100.0** (250)	**100.0** (250)
P2G	0.0 (250)	0.0 (250)	4.0 (250)
A2G	7.2 (250)	1.2 (250)	0.4 (250)
HM	0.8 (250)	0.0 (250)	10.8 (250)
M_{ps}	0.4 (250)	1.2 (250)	2.8 (250)
S_{ps}	0.0 (250)	1.2 (250)	0.8 (250)
$MS1_{ps}$	**100.0** (250)	**93.2** (250)	**100.0** (250)
$MS2_{ps}$	**100.0** (250)	**98.8** (250)	**100.0** (250)
M_{pL}	0.8 (250)	0.4 (250)	0.4 (250)
S_{pL}	0.4 (250)	1.2 (250)	1.2 (250)
$MS1_{pL}$	**100.0** (250)	**89.2** (250)	**100.0** (250)
$MS2_{pL}$	**100.0** (250)	**98.0** (250)	**100.0** (250)
M_t	0.4 (250)	0.4 (250)	1.6 (250)
S_t	5.2 (250)	5.6 (250)	3.2 (250)
$MS1_t$	**100.0** (250)	**100.0** (250)	**100.0** (250)
$MS2_t$	**100.0** (250)	**100.0** (250)	**100.0** (250)
$SD1_t$	15.2 (250)	0.0 (250)	0.0 (250)

Table 3.8: Test of positive association (conservative variance estimate), split by state, 2 informative movements (animal captured at least 5 times) , percentage of significant results (number of applicable tests), high percentage of significant results in bold (> 50%)

scenario	**% (N)**		
	State 1	**State 2**	**State 3**
M	0.4 (250)	0.0 (250)	1.2 (250)
S	0.8 (250)	0.4 (250)	0.0 (250)
MO	0.0 (250)	0.4 (250)	0.8 (250)
P	4.0 (250)	0.0 (250)	0.4 (250)
A	0.0 (250)	0.0 (250)	0.4 (250)
SD1	0.4 (250)	0.4 (250)	0.0 (250)
SD3	0.4 (250)	0.0 (250)	0.0 (250)
Mem1	**97.6** (250)	**100.0** (250)	**100.0** (250)
Mem2	**85.2** (250)	**99.2** (250)	**99.2** (250)
MS1	**100.0** (250)	**100.0** (250)	**100.0** (250)
MS2	**100.0** (250)	**100.0** (250)	**100.0** (250)
P2G	6.4 (250)	2.8 (250)	3.2 (250)
A2G	4.8 (250)	0.0 (250)	0.0 (250)
HM	0.0 (250)	0.0 (250)	0.4 (250)
M_{ps}	0.0 (250)	0.0 (250)	1.2 (250)
S_{ps}	0.0 (250)	0.4 (250)	2.0 (250)
$MS1_{ps}$	**100.0** (250)	**58.8** (250)	**100.0** (250)
$MS2_{ps}$	**100.0** (250)	**94.8** (250)	**100.0** (250)
M_{pL}	0.0 (250)	0.4 (250)	0.8 (250)
S_{pL}	0.4 (250)	0.4 (250)	0.4 (250)
$MS1_{pL}$	**98.4** (250)	**75.6** (250)	**99.2** (250)
$MS2_{pL}$	**98.8** (250)	**92.4** (250)	**98.8** (250)
M_t	0.0 (250)	0.4 (250)	0.0 (250)
S_t	0.8 (250)	0.8 (250)	0.4 (250)
$MS1_t$	**100.0** (250)	**100.0** (250)	**100.0** (250)
$MS2_t$	**100.0** (250)	**98.4** (250)	**100.0** (250)
$SD1_t$	0.4 (250)	0.0 (250)	0.0 (250)

Table 3.9: Summarised conservative test of positive association based on 1 informative movement (denoted by 1 IM, animals captured at least 3 times) and 2 informative movements (denoted by 2 IM, animals captured at least 5 times),percentage of significant results (number of applicable tests), high percentage of significant results in bold (> 50%)

Scenario	**% C, 1IM (N)**	**% C, 2IM (N)**
M	7.2 (250)	0.8 (250)
S	4.0 (250)	0.8 (250)
MO	0.0 (250)	0.0 (250)
P	0.8 (250)	0.4 (250)
A	1.2 (250)	0.0 (250)
SD1	3.2 (250)	0.4 (250)
SD3	0.4 (250)	0.4 (250)
Mem1	**100.0** (250)	**100.0** (250)
Mem2	**100.0** (250)	**100.0** (250)
MS1	**100.0** (250)	**100.0** (250)
MS2	**100.0** (250)	**100.0** (250)
P2G	0.4 (250)	17.2 (250)
A2G	5.6 (250)	2.4 (250)
HM	5.2 (250)	0.0 (250)
M_{ps}	0.4 (250)	0.0 (250)
S_{ps}	0.0 (250)	0.8 (250)
$MS1_{ps}$	**100.0** (250)	**100.0** (250)
$MS2_{ps}$	**100.0** (250)	**100.0** (250)
M_{pL}	1.2 (250)	0.8 (250)
S_{pL}	0.4 (250)	0.8 (250)
$MS1_{pL}$	**100.0** (250)	**100.0** (250)
$MS2_{pL}$	**100.0** (250)	**100.0** (250)
M_t	0.4 (250)	0.4 (250)
S_t	14.0 (250)	0.4 (250)
$MS1_t$	**100.0** (250)	**100.0** (250)
$MS2_t$	**100.0** (250)	**100.0** (250)
$SD1_t$	2.4 (250)	0.4 (250)

Existing memory tests

We investigated the tests currently used to detect memory, in order to assess whether they were actually specific to memory. For these tests, time-dependent scenarios were not considered since the tests are based on independent components by state at each occasion. Test WBWA was coded by adapting MATLAB code provided by R. Choquet; we used the Kappa function from R package `vcd` to obtain the kappa estimate, its asymptotic standard error and the resulting z-statistic (Meyer et al., 2016). We used both a one-sided test corresponding to $\kappa > 0$: more agreement than expected by chance (which is the case for memory) and a two-sided test which also adds to the alternative $\kappa < 0$ (less agreement than expected by chance, which would correspond to animals avoiding the site where they were last seen) (see for example Everitt, 1992, p.148).

Table 3.10 showed that the global WBWA test (formed by summing the WBWA tests by occasion and state) reacts strongly not only to memory, but also to the existence of a mover-stayer structure, heterogeneity in preferences or avoidance as well as heterogeneity in movement, with close to 100% of significant results for all these situations. In Tables 3.11 to 3.14, we show the results obtained using test WBWA split by state and occasion. As expected, the split WBWA test shows similar reactions as the global test, though not always as strong or for all states (e.g. see Table 3.13).

The results obtained using the kappa statistic by occasion and state are displayed, for informative purposes, in Tables 3.15 to 3.18; they are very similar to the results obtained with Test WBWA: the test reacts strongly to both memory and a mover-stayer structure; it is also sensitive to 2 groups with different preferences and to heterogeneity in movement. Due to the similarities between the results from Test WBWA and Cohen's kappa, we stopped here and did not deem it necessary to pursue this route further.

Table 3.10: Global Test WBWA, percentage of significant results (number of applicable tests), high percentage of significant results in bold (> 50%)

Scenario	**% (N)**
M	3.6 (250)
S	4.8 (250)
MO	4.0 (250)
P	3.2 (250)
A	6.0 (250)
SD1	3.2 (250)
SD3	6.4 (250)
Mem1	**100.0** (250)
Mem2	**100.0** (250)
MS1	**100.0** (250)
MS2	**100.0** (250)
P2G	**100.0** (250)
A2G	**98.4** (250)
HM	**96.4** (250)
M_{ps}	5.6 (250)
S_{ps}	1.2 (250)
$MS1_{ps}$	**100.0** (250)
$MS2_{ps}$	**100.0** (250)
M_{pL}	4.0 (250)
S_{pL}	1.6 (250)
$MS1_{pL}$	**100.0** (250)
$MS2_{pL}$	**100.0** (250)

Table 3.11: Test WBWA by occasion and state, homogeneous scenarios, percentage of significant results (number of applicable tests), high percentage of significant results in bold (> 50%)

Scenario	Occasion	S1 (N)	S2 (N)	S3 (N)
M	2	5.2 (250)	6.8 (250)	5.2 (250)
M	3	5.2 (250)	6.0 (250)	5.2 (250)
M	4	5.6 (250)	4.8 (250)	6.8 (250)
M	5	4.8 (250)	4.0 (250)	3.2 (250)
M	6	3.2 (250)	3.2 (250)	4.4 (250)
M	7	3.2 (250)	4.4 (250)	4.8 (250)
M	8	4.0 (250)	6.0 (250)	5.2 (250)
M	9	6.4 (250)	5.2 (250)	6.4 (250)
S	2	2.0 (250)	1.6 (250)	4.0 (250)
S	3	1.6 (250)	5.2 (250)	4.8 (250)
S	4	3.6 (250)	5.2 (250)	6.0 (250)
S	5	3.2 (250)	6.4 (250)	4.4 (250)
S	6	5.2 (250)	1.2 (250)	5.6 (250)
S	7	2.4 (250)	5.6 (250)	5.6 (250)
S	8	4.0 (250)	6.4 (250)	6.0 (250)
S	9	4.0 (250)	6.8 (250)	6.4 (250)
MO	2	5.6 (250)	6.0 (250)	5.2 (250)
MO	3	4.0 (250)	3.2 (250)	2.0 (250)
MO	4	2.0 (250)	4.8 (250)	4.0 (250)
MO	5	6.0 (250)	4.8 (250)	1.6 (250)
MO	6	6.8 (250)	4.8 (250)	3.2 (250)
MO	7	4.0 (250)	2.8 (250)	4.4 (250)
MO	8	4.4 (250)	4.0 (250)	5.2 (250)
MO	9	4.4 (250)	5.2 (250)	6.0 (250)
P	2	2.0 (250)	5.6 (250)	3.2 (250)
P	3	6.0 (250)	2.4 (250)	3.2 (250)
P	4	4.8 (250)	7.6 (250)	2.0 (250)
P	5	4.0 (250)	2.8 (250)	4.8 (250)
P	6	5.6 (250)	5.2 (250)	5.2 (250)
P	7	3.6 (250)	2.4 (250)	2.8 (250)
P	8	4.0 (250)	6.4 (250)	4.8 (250)
P	9	5.6 (250)	5.2 (250)	2.4 (250)
A	2	6.8 (250)	2.8 (250)	5.6 (250)
A	3	6.8 (250)	6.0 (250)	4.0 (250)
A	4	4.4 (250)	6.8 (250)	9.2 (250)
A	5	6.8 (250)	3.6 (250)	5.6 (250)
A	6	4.8 (250)	5.6 (250)	7.2 (250)
A	7	5.6 (250)	3.2 (250)	4.0 (250)
A	8	6.0 (250)	5.6 (250)	6.4 (250)
A	9	4.0 (250)	6.0 (250)	4.0 (250)
SD1	2	6.0 (250)	4.8 (250)	4.0 (250)
SD1	3	3.2 (250)	6.0 (250)	4.0 (250)
SD1	4	7.6 (250)	3.2 (250)	2.8 (250)
SD1	5	8.8 (250)	4.4 (250)	5.6 (250)
SD1	6	2.8 (250)	6.0 (250)	4.4 (250)
SD1	7	3.6 (250)	2.4 (250)	5.2 (250)
SD1	8	3.6 (250)	5.6 (250)	5.2 (250)
SD1	9	3.6 (250)	4.8 (250)	6.8 (250)
SD3	2	1.6 (250)	7.2 (250)	4.4 (250)
SD3	3	4.8 (250)	8.8 (250)	3.2 (250)
SD3	4	6.8 (250)	4.4 (250)	6.0 (250)
SD3	5	4.0 (250)	3.2 (250)	4.4 (250)
SD3	6	4.4 (250)	4.4 (250)	5.2 (250)
SD3	7	3.6 (250)	5.6 (250)	4.0 (250)
SD3	8	5.2 (250)	4.8 (250)	3.2 (250)
SD3	9	2.8 (250)	5.2 (250)	6.0 (250)

Table 3.12: Test WBWA by occasion, memory scenarios, percentage of significant results (number of applicable tests), high percentage of significant results in bold (> 50%)

Scenario	Occasion	S1 (N)	S2 (N)	S3 (N)
Mem1	2	**99.2** (250)	**99.6** (250)	**100** (250)
Mem1	3	**100.0** (250)	**100.0** (250)	**100** (250)
Mem1	4	**100.0** (250)	**100.0** (250)	**100** (250)
Mem1	5	**100.0** (250)	**100.0** (250)	**100** (250)
Mem1	6	**100.0** (250)	**100.0** (250)	**100** (250)
Mem1	7	**100.0** (250)	**100.0** (250)	**100** (250)
Mem1	8	**100.0** (250)	**100.0** (250)	**100** (250)
Mem1	9	**100.0** (250)	**100.0** (250)	**100** (250)
Mem2	2	**95.2** (250)	**98.0** (250)	**99.2** (250)
Mem2	3	**100.0** (250)	**100.0** (250)	**100** (250)
Mem2	4	**100.0** (250)	**100.0** (250)	**100** (250)
Mem2	5	**100.0** (250)	**100.0** (250)	**100** (250)
Mem2	6	**100.0** (250)	**100.0** (250)	**100** (250)
Mem2	7	**100.0** (250)	**100.0** (250)	**100** (250)
Mem2	8	**100.0** (250)	**100.0** (250)	**100** (250)
Mem2	9	**100.0** (250)	**100.0** (250)	**100.0** (250)

Table 3.13: Test WBWA by occasion, heterogeneous scenarios, part 1, percentage of significant results (number of applicable tests), high percentage of significant results in bold (> 50%)

Scenario	Occasion	S1 (N)	S2 (N)	S3 (N)
MS1	2	**69.6** (250)	28.8 (250)	**86.4** (250)
MS1	3	**92.8** (250)	42.0 (250)	**100.0** (250)
MS1	4	**98.8** (250)	**56.8** (250)	**100.0** (250)
MS1	5	**99.6** (250)	**64.8** (250)	**100.0** (250)
MS1	6	**100.0** (250)	**77.6** (250)	**100.0** (250)
MS1	7	**100.0** (250)	**75.2** (250)	**100.0** (250)
MS1	8	**100.0** (250)	**77.2** (250)	**100.0** (250)
MS1	9	**100.0** (250)	**77.2** (250)	**100.0** (250)
MS2	2	**82.8** (250)	49.2 (250)	**92.0** (250)
MS2	3	**98.0** (250)	**64.4** (250)	**98.4** (250)
MS2	4	**99.2** (250)	**79.6** (250)	**100.0** (250)
MS2	5	**100.0** (250)	**86.0** (250)	**100.0** (250)
MS2	6	**100.0** (250)	**90.4** (250)	**100.0** (250)
MS2	7	**100.0** (250)	**92.4** (250)	**100.0** (250)
MS2	8	**100.0** (250)	**95.6** (250)	**100.0** (250)
MS2	9	**100.0** (250)	**95.6** (250)	**100.0** (250)
P2G	2	8.0 (250)	8.8 (250)	12.0 (250)
P2G	3	10.0 (250)	8.4 (250)	5.2 (250)
P2G	4	31.2 (250)	29.6 (250)	10.0 (250)
P2G	5	**55.6** (250)	**54.8** (250)	22.0 (250)
P2G	6	**75.6** (250)	**74.8** (250)	31.6 (250)
P2G	7	**84.4** (250)	**90.0** (250)	41.6 (250)
P2G	8	**91.6** (250)	**92.4** (250)	**51.2** (250)
P2G	9	**92.4** (250)	**96.8** (250)	**59.6** (250)
A2G	2	9.2 (250)	7.6 (250)	3.6 (250)
A2G	3	12.0 (250)	10.8 (250)	6.4 (250)
A2G	4	16.8 (250)	15.2 (250)	9.6 (250)
A2G	5	26.4 (250)	20.0 (250)	10.8 (250)
A2G	6	32.0 (250)	29.2 (250)	18.4 (250)
A2G	7	36.0 (250)	40.8 (250)	21.2 (250)
A2G	8	45.2 (250)	45.6 (250)	22.8 (250)
A2G	9	49.6 (250)	45.6 (250)	20.0 (250)
HM	2	11.2 (250)	7.2 (250)	7.2 (250)
HM	3	28.4 (250)	6.4 (250)	5.6 (250)
HM	4	41.2 (250)	7.2 (250)	5.6 (250)
HM	5	48.4 (250)	8.8 (250)	7.2 (250)
HM	6	**58.4** (250)	8.4 (250)	6.8 (250)
HM	7	**66.4** (250)	8.0 (250)	5.6 (250)
HM	8	**72.0** (250)	9.2 (250)	6.8 (250)
HM	9	**74.4** (250)	5.6 (250)	9.2 (250)

Table 3.14: Test WBWA by occasion, heterogeneous scenarios, part 2, percentage of significant results (number of applicable tests), high percentage of significant results in bold (> 50%)

Scenario	Occasion	S1 (N)	S2 (N)	S3 (N)
M_{ps}	2	4.4 (250)	3.6 (250)	5.2 (250)
M_{ps}	3	4.8 (250)	4.0 (250)	1.2 (250)
M_{ps}	4	4.8 (250)	5.6 (250)	6.0 (250)
M_{ps}	5	6.4 (250)	6.4 (250)	5.2 (250)
M_{ps}	6	5.2 (250)	6.8 (250)	4.4 (250)
M_{ps}	7	5.2 (250)	2.8 (250)	4.0 (250)
M_{ps}	8	3.2 (250)	3.2 (250)	5.6 (250)
M_{ps}	9	3.6 (250)	4.8 (250)	5.6 (250)
S_{ps}	2	1.6 (250)	1.2 (250)	3.6 (250)
S_{ps}	3	5.2 (250)	3.2 (250)	3.6 (250)
S_{ps}	4	5.2 (250)	4.8 (250)	6.0 (250)
S_{ps}	5	4.4 (250)	6.4 (250)	6.0 (250)
S_{ps}	6	4.0 (250)	6.4 (250)	4.0 (250)
S_{ps}	7	4.4 (250)	5.2 (250)	4.8 (250)
S_{ps}	8	6.4 (250)	2.8 (250)	2.8 (250)
S_{ps}	9	4.0 (250)	4.4 (250)	1.6 (250)
$MS1_{ps}$	2	45.6 (250)	13.2 (250)	**60.4** (250)
$MS1_{ps}$	3	**70.8** (250)	15.6 (250)	**83.2** (250)
$MS1_{ps}$	4	**83.6** (250)	20.0 (250)	**90.8** (250)
$MS1_{ps}$	5	**87.2** (250)	22.4 (250)	**96.8** (250)
$MS1_{ps}$	6	**91.6** (250)	28.4 (250)	**99.2** (250)
$MS1_{ps}$	7	**94.8** (250)	30.0 (250)	**100.0** (250)
$MS1_{ps}$	8	**98.0** (250)	29.2 (250)	**100.0** (250)
$MS1_{ps}$	9	**100.0** (250)	19.6 (250)	**100.0** (250)
$MS2_{ps}$	2	**72.8** (250)	21.2 (250)	**74** (250)
$MS2_{ps}$	3	**84.8** (250)	25.2 (250)	**90.4** (250)
$MS2_{ps}$	4	**88.4** (250)	24.0 (250)	**95.2** (250)
$MS2_{ps}$	5	**97.6** (250)	34.8 (250)	**97.6** (250)
$MS2_{ps}$	6	**98.8** (250)	36.0 (250)	**100.0** (250)
$MS2_{ps}$	7	**97.6** (250)	37.2 (250)	**99.6** (250)
$MS2_{ps}$	8	**100.0** (250)	32.4 (250)	**100.0** (250)
$MS2_{ps}$	9	**99.6** (250)	39.2 (250)	**100.0** (250)
M_{pL}	2	6.4 (250)	4.0 (250)	5.6 (250)
M_{pL}	3	6.0 (250)	2.4 (250)	4.8 (250)
M_{pL}	4	6.0 (250)	5.6 (250)	6.4 (250)
M_{pL}	5	4.0 (250)	7.2 (250)	5.6 (250)
M_{pL}	6	4.0 (250)	5.2 (250)	4.8 (250)
M_{pL}	7	4.0 (250)	4.8 (250)	7.2 (250)
M_{pL}	8	4.4 (250)	5.2 (250)	5.6 (250)
M_{pL}	9	3.2 (250)	5.2 (250)	5.2 (250)
S_{pL}	2	0.8 (250)	2.4 (250)	3.2 (250)
S_{pL}	3	3.2 (250)	6.4 (250)	3.2 (250)
S_{pL}	4	3.6 (250)	7.6 (250)	3.6 (250)
S_{pL}	5	4.4 (250)	2.4 (250)	3.6 (250)
S_{pL}	6	4.8 (250)	6.0 (250)	5.2 (250)
S_{pL}	7	4.4 (250)	6.4 (250)	4.8 (250)
S_{pL}	8	5.2 (250)	4.8 (250)	2.8 (250)
S_{pL}	9	2 (250)	2.4 (250)	6.4 (250)
$MS1_{pL}$	2	33.6 (250)	11.6 (250)	32.8 (250)
$MS1_{pL}$	3	43.6 (250)	13.6 (250)	45.6 (250)
$MS1_{pL}$	4	**61.2** (250)	14.0 (250)	**67.6** (250)
$MS1_{pL}$	5	**66.4** (250)	14.4 (250)	**74.8** (250)
$MS1_{pL}$	6	**73.6** (250)	19.6 (250)	**76.0** (250)
$MS1_{pL}$	7	**82.4** (250)	18.4 (250)	**82.0** (250)
$MS1_{pL}$	8	**84.4** (250)	20.4 (250)	**88.8** (250)
$MS1_{pL}$	9	**80.4** (250)	18.8 (250)	**88.8** (250)
$MS2_{pL}$	2	41.2 (250)	20.8 (250)	46.8 (250)
$MS2_{pL}$	3	**61.6** (250)	19.2 (250)	**52.0** (250)
$MS2_{pL}$	4	**71.6** (250)	28.0 (250)	**62.4** (250)
$MS2_{pL}$	5	**73.2** (250)	28.8 (250)	**72.4** (250)
$MS2_{pL}$	6	**82.0** (250)	29.6 (250)	**77.6** (250)
$MS2_{pL}$	7	**83.6** (250)	31.6 (250)	**81.6** (250)
$MS2_{pL}$	8	**89.6** (250)	31.6 (250)	**87.2** (250)
$MS2_{pL}$	9	**86.4** (250)	42.4 (250)	**89.6** (250)

Table 3.15: Kappa test by occasion, homogeneous scenarios, percentage of significant results (number of applicable tests), high percentage of significant results in bold (> 50%)

Scenario	**Occasion**	**1-sided** ($\kappa > 0$)			**2-sided** ($\kappa \neq 0$)		
		S1 (N)	S2 (N)	S3 (N)	S1 (N)	S2 (N)	S3 (N)
M	2	4.4 (250)	3.6 (250)	5.2 (250)	7.6 (250)	5.2 (250)	5.2 (250)
M	3	7.6 (250)	3.6 (250)	6 (250)	7.2 (250)	5.2 (250)	6.0 (250)
M	4	6.0 (250)	5.2 (250)	6.4 (250)	7.6 (250)	6.8 (250)	9.2 (250)
M	5	4.0 (250)	4.0 (250)	4.8 (250)	4.0 (250)	4.4 (250)	4.0 (250)
M	6	4.4 (250)	3.6 (250)	4.4 (250)	2.8 (250)	4.4 (250)	6.0 (250)
M	7	4.8 (250)	6.0 (250)	2.4 (250)	5.6 (250)	4.4 (250)	3.2 (250)
M	8	6.8 (250)	4.0 (250)	3.2 (250)	5.6 (250)	3.6 (250)	4.0 (250)
M	9	8.0 (250)	4.4 (250)	3.6 (250)	6.8 (250)	3.2 (250)	5.6 (250)
S	2	2.0 (250)	2.0 (250)	2.0 (250)	30.0 (250)	24.0 (250)	25.2 (250)
S	3	1.2 (250)	2.8 (250)	2.0 (250)	14.8 (250)	12.0 (250)	17.6 (250)
S	4	3.6 (250)	2.8 (250)	5.2 (250)	8.8 (250)	9.6 (250)	14.4 (250)
S	5	0.8 (250)	4.8 (250)	2.8 (250)	7.6 (250)	9.6 (250)	11.6 (250)
S	6	3.2 (250)	3.6 (250)	3.2 (250)	15.6 (250)	8.0 (250)	10.8 (250)
S	7	2.4 (250)	3.6 (250)	3.6 (250)	10.4 (250)	6.0 (250)	6 (250)
S	8	3.2 (250)	2.4 (250)	5.2 (250)	7.6 (250)	5.6 (250)	8.4 (250)
S	9	0.4 (250)	3.6 (250)	3.2 (250)	10.4 (250)	5.2 (250)	11.2 (250)
MO	2	6.4 (250)	5.6 (250)	7.6 (250)	6.8 (250)	6.0 (250)	5.6 (250)
MO	3	2.8 (250)	4.8 (250)	5.2 (250)	6.4 (250)	4.8 (250)	2.8 (250)
MO	4	2.4 (250)	5.2 (250)	5.6 (250)	4.8 (250)	4.8 (250)	4.4 (250)
MO	5	6.8 (250)	7.2 (250)	4.4 (250)	5.2 (250)	5.6 (250)	4.4 (250)
MO	6	5.2 (250)	5.6 (250)	3.2 (250)	5.6 (250)	6.4 (250)	4.0 (250)
MO	7	4.0 (250)	4.0 (250)	6.0 (250)	2.8 (250)	3.2 (250)	6.8 (250)
MO	8	4.8 (250)	4.4 (250)	2.8 (250)	6.0 (250)	5.6 (250)	4.8 (250)
MO	9	5.6 (250)	6.0 (250)	3.6 (250)	6.0 (250)	8 (250)	1.6 (250)
P	2	3.6 (250)	4.8 (250)	5.2 (250)	2.4 (250)	5.2 (250)	3.2 (250)
P	3	5.6 (250)	2.0 (250)	3.2 (250)	4.4 (250)	6.8 (250)	4.0 (250)
P	4	5.2 (250)	3.6 (250)	4.8 (250)	6.8 (250)	6.0 (250)	4.8 (250)
P	5	6.4 (250)	2.8 (250)	4.8 (250)	7.6 (250)	6.4 (250)	6.4 (250)
P	6	3.6 (250)	2.8 (250)	7.6 (250)	6.0 (250)	8.4 (250)	10.0 (250)
P	7	5.2 (250)	2.4 (250)	4.0 (250)	7.6 (250)	6.8 (250)	3.6 (250)
P	8	2.8 (250)	4.0 (250)	3.2 (250)	8.4 (250)	8.4 (250)	3.6 (250)
P	9	4.4 (250)	3.2 (250)	4.8 (250)	12.8 (250)	8.0 (250)	7.6 (250)
A	2	4.8 (250)	4.0 (250)	5.6 (250)	6.8 (250)	7.2 (250)	6.4 (250)
A	3	2.8 (250)	3.2 (250)	6.8 (250)	4.0 (250)	6.0 (250)	6.0 (250)
A	4	2.4 (250)	5.6 (250)	3.2 (250)	4.4 (250)	6.0 (250)	7.2 (250)
A	5	4.0 (250)	4.4 (250)	4.4 (250)	5.2 (250)	5.6 (250)	7.6 (250)
A	6	2.4 (250)	6.8 (250)	7.6 (250)	1.6 (250)	6.4 (250)	6.0 (250)
A	7	4.0 (250)	2.4 (250)	5.2 (250)	4.8 (250)	4.4 (250)	6.8 (250)
A	8	8.0 (250)	2.0 (250)	2.0 (250)	6.4 (250)	4.8 (250)	4.4 (250)
A	9	3.6 (250)	5.2 (250)	6.0 (250)	4.0 (250)	5.2 (250)	5.6 (250)
SD1	2	5.2 (250)	2.8 (250)	3.6 (250)	6.4 (250)	6.4 (250)	6.0 (250)
SD1	3	2.8 (250)	4.0 (250)	5.6 (250)	4.0 (250)	5.2 (250)	7.6 (250)
SD1	4	7.2 (250)	2.4 (250)	6.8 (250)	9.6 (250)	4.4 (250)	5.6 (250)
SD1	5	6.8 (250)	5.6 (250)	4.8 (250)	4.0 (250)	5.2 (250)	5.2 (250)
SD1	6	4.4 (250)	1.6 (250)	4.0 (250)	6.0 (250)	2.4 (250)	3.6 (250)
SD1	7	4.0 (250)	2.8 (250)	4.0 (250)	5.2 (250)	7.6 (250)	4.8 (250)
SD1	8	5.2 (250)	6.0 (250)	1.6 (250)	5.2 (250)	6.4 (250)	5.6 (250)
SD1	9	6.8 (250)	3.6 (250)	4.4 (250)	4.4 (250)	4.0 (250)	4.4 (250)
SD3	2	2.0 (250)	7.6 (250)	4.0 (250)	26.0 (250)	9.2 (250)	5.6 (250)
SD3	3	3.2 (250)	4.8 (250)	4.0 (250)	10.4 (250)	6.0 (250)	2.8 (250)
SD3	4	5.2 (250)	7.2 (250)	6.0 (250)	6.8 (250)	6.4 (250)	6.0 (250)
SD3	5	3.2 (250)	2.4 (250)	6.0 (250)	4.0 (250)	5.2 (250)	6.0 (250)
SD3	6	4.8 (250)	4.4 (250)	6.0 (250)	6.0 (250)	4.4 (250)	5.6 (250)
SD3	7	4.8 (250)	4.0 (250)	6.0 (250)	4.8 (250)	4.8 (250)	4.8 (250)
SD3	8	5.2 (250)	2.0 (250)	4.4 (250)	7.2 (250)	5.6 (250)	3.2 (250)
SD3	9	1.2 (250)	4.8 (250)	7.6 (250)	5.2 (250)	4.8 (250)	6.0 (250)

Table 3.16: Kappa test by occasion, memory scenarios, percentage of significant results (number of applicable tests), high percentage of significant results in bold ($> 50\%$)

Scenario	**Occasion**	**1-sided** ($\kappa > 0$)			**2-sided** ($\kappa \neq 0$)		
		S1 (N)	S2 (N)	S3 (N)	S1 (N)	S2 (N)	S3 (N)
Mem1	2	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**99.6** (250)	**100.0** (250)
Mem1	3	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)
Mem1	4	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)
Mem1	5	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)
Mem1	6	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)
Mem1	7	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)
Mem1	8	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)
Mem1	9	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)
Mem2	2	**99.2** (250)	**99.6** (250)	**100.0** (250)	**98.4** (250)	**99.6** (250)	**100.0** (250)
Mem2	3	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)
Mem2	4	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)
Mem2	5	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)
Mem2	6	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)
Mem2	7	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)
Mem2	8	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)
Mem2	9	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)	**100.0** (250)

Conclusions

Our simulation results show that the significance of the WBWA test, currently used as a test for memory, could actually be indicative of animals with different preferences, with heterogeneous movement patterns, or a mover-stayer structure. Based on the simulation scenarios considered, the test of positive association reacts to a smaller subset of situations: mainly memory and mover-stayer. Therefore in Section 3.5, we attempt to construct adaptations of both Test WBWA and the positive association test that would allow us to identify specifically either the existence of a mover-stayer structure or the presence of short-term memory.

Table 3.17: Kappa test by occasion, heterogeneous scenarios, part 1, percentage of significant results (number of applicable tests), high percentage of significant results in bold (> 50%)

Scenario	**Occasion**	**1-sided** ($\kappa > 0$)			**2-sided** ($\kappa \neq 0$)		
		S1 (N)	S2 (N)	S3 (N)	S1 (N)	S2 (N)	S3 (N)
MS1	2	74 (250)	42.8 (250)	86.4 (250)	64.8 (250)	29.2 (250)	78.4 (250)
MS1	3	**93.2** (250)	**56.4** (250)	**98.0** (250)	**88.8** (250)	43.2 (250)	**95.6** (250)
MS1	4	**98.8** (250)	**69.2** (250)	**100.0** (250)	**97.2** (250)	**58.4** (250)	**99.6** (250)
MS1	5	**99.6** (250)	**73.2** (250)	**100** (250)	**99.2** (250)	**60.8** (250)	**99.2** (250)
MS1	6	**100.0** (250)	**78.8** (250)	**100.0** (250)	**100.0** (250)	**67.2** (250)	**100.0** (250)
MS1	7	**100.0** (250)	**78.4** (250)	**100.0** (250)	**100.0** (250)	**65.2** (250)	**100.0** (250)
MS1	8	**100.0** (250)	**84.4** (250)	**100** (250)	**100** (250)	**76.4** (250)	**100** (250)
MS1	9	**100** (250)	**86.8** (250)	**100.0** (250)	**100.0** (250)	**77.2** (250)	**100.0** (250)
MS2	2	**83.6** (250)	**57.2** (250)	**93.6** (250)	**76.8** (250)	44.0 (250)	**88.4** (250)
MS2	3	**99.2** (250)	**80.0** (250)	**99.6** (250)	**97.2** (250)	**70.4** (250)	**98.8** (250)
MS2	4	**99.6** (250)	**89.6** (250)	**100.0** (250)	**99.6** (250)	**80.4** (250)	**100.0** (250)
MS2	5	**100.0** (250)	**90.4** (250)	**100.0** (250)	**100.0** (250)	**89.2** (250)	**100.0** (250)
MS2	6	**100.0** (250)	**95.2** (250)	**100.0** (250)	**100.0** (250)	**91.6** (250)	**100.0** (250)
MS2	7	**100.0** (250)	**97.6** (250)	**100.0** (250)	**100.0** (250)	**94.8** (250)	**100.0** (250)
MS2	8	**100.0** (250)	**96.8** (250)	**100.0** (250)	**100.0** (250)	**94.8** (250)	**100.0** (250)
MS2	9	**100.0** (250)	**97.6** (250)	**100.0** (250)	**100.0** (250)	**96.0** (250)	**100.0** (250)
P2G	2	2.0 (250)	0.4 (250)	0.0 (250)	7.2 (250)	11.6 (250)	22.8 (250)
P2G	3	16.0 (250)	9.2 (250)	4.4 (250)	8.0 (250)	6.8 (250)	8.4 (250)
P2G	4	39.6 (250)	30.4 (250)	20.0 (250)	27.6 (250)	22.0 (250)	14.0 (250)
P2G	5	**60.8** (250)	**53.6** (250)	**40.0** (250)	**46.8** (250)	**42.0** (250)	29.6 (250)
P2G	6	**79.2** (250)	**77.6** (250)	**54.4** (250)	**67.2** (250)	**64.4** (250)	44 (250)
P2G	7	**88.4** (250)	**86.0** (250)	**62.4** (250)	**80.0** (250)	**74.8** (250)	**50.0** (250)
P2G	8	**93.2** (250)	**92.4** (250)	**69.6** (250)	**88.4** (250)	**83.6** (250)	**59.2** (250)
P2G	9	**94** (250)	**96** (250)	**79.2** (250)	**87.6** (250)	**86** (250)	**69.6** (250)
A2G	2	9.2 (250)	4.8 (250)	3.2 (250)	8.8 (250)	6.4 (250)	8.4 (250)
A2G	3	13.6 (250)	9.2 (250)	8.0 (250)	8.4 (250)	6.4 (250)	4.8 (250)
A2G	4	17.6 (250)	15.2 (250)	13.6 (250)	11.6 (250)	10 (250)	8.4 (250)
A2G	5	30.8 (250)	19.2 (250)	12.0 (250)	20.8 (250)	10.4 (250)	8.8 (250)
A2G	6	36.4 (250)	21.2 (250)	21.2 (250)	24.4 (250)	14 (250)	13.2 (250)
A2G	7	37.6 (250)	27.2 (250)	30.4 (250)	27.2 (250)	18.4 (250)	18.0 (250)
A2G	8	40.8 (250)	26.0 (250)	25.2 (250)	30.4 (250)	15.6 (250)	15.2 (250)
A2G	9	48 (250)	32.8 (250)	19.6 (250)	38.4 (250)	20.8 (250)	12.4 (250)
HM	2	15.6 (250)	2.4 (250)	1.2 (250)	9.2 (250)	7.2 (250)	8 (250)
HM	3	30.8 (250)	4.4 (250)	4.8 (250)	19.6 (250)	3.6 (250)	7.2 (250)
HM	4	45.6 (250)	6.4 (250)	8.4 (250)	30.0 (250)	4.8 (250)	6.4 (250)
HM	5	49.2 (250)	7.6 (250)	9.2 (250)	38 (250)	6.4 (250)	8.4 (250)
HM	6	**56.8** (250)	7.2 (250)	12 (250)	42.4 (250)	6.8 (250)	9.6 (250)
HM	7	**62.8** (250)	4.8 (250)	12.4 (250)	**52.4** (250)	3.6 (250)	8.4 (250)
HM	8	**72.4** (250)	8.8 (250)	15.6 (250)	**59.6** (250)	8.8 (250)	10.0 (250)
HM	9	**73.6** (250)	6.0 (250)	17.6 (250)	**59.6** (250)	4.0 (250)	13.2 (250)

Table 3.18: Kappa test by occasion, heterogeneous scenarios, part 2, percentage of significant results (number of applicable tests), high percentage of significant results in bold (> 50%)

Scenario	Occasion	1-sided ($\kappa > 0$)			2-sided ($\kappa \neq 0$)		
		S1 (N)	S2 (N)	S3 (N)	S1 (N)	S2 (N)	S3 (N)
M_{ps}	2	4.4 (250)	5.6 (250)	4.0 (250)	5.6 (250)	8.4 (250)	6.0 (250)
M_{ps}	3	4.8 (250)	4.4 (250)	3.6 (250)	4.8 (250)	6.8 (250)	6.0 (250)
M_{ps}	4	5.2 (250)	4.8 (250)	3.2 (250)	6.8 (250)	6.8 (250)	4.8 (250)
M_{ps}	5	2.8 (250)	5.6 (250)	3.6 (250)	6.4 (250)	6.4 (250)	6.4 (250)
M_{ps}	6	3.2 (250)	5.6 (250)	4.8 (250)	5.2 (250)	6.0 (250)	5.6 (250)
M_{ps}	7	8.0 (250)	4.4 (250)	3.6 (250)	8.8 (250)	5.6 (250)	4.8 (250)
M_{ps}	8	2.4 (250)	4.0 (250)	4.8 (250)	5.6 (250)	4.8 (250)	4.8 (250)
M_{ps}	9	6.0 (250)	4.0 (250)	2.4 (250)	4.4 (250)	4.8 (250)	2.4 (250)
S_{ps}	2	1.2 (250)	2.8 (250)	3.6 (250)	31.6 (250)	9.2 (250)	28.0 (250)
S_{ps}	3	2.8 (250)	3.6 (250)	1.6 (250)	20.0 (250)	12.8 (250)	10.4 (250)
S_{ps}	4	4.0 (250)	2.8 (250)	2.8 (250)	15.6 (250)	8.0 (250)	9.6 (250)
S_{ps}	5	2.4 (250)	3.6 (250)	3.2 (250)	10.0 (250)	6.4 (250)	6.4 (250)
S_{ps}	6	1.6 (250)	4.0 (250)	2.0 (250)	15.2 (250)	5.6 (250)	13.2 (250)
S_{ps}	7	2.4 (250)	5.2 (250)	3.6 (250)	11.2 (250)	8.0 (250)	8.8 (250)
S_{ps}	8	2.8 (250)	2.0 (250)	1.6 (250)	10.8 (250)	7.2 (250)	6.8 (250)
S_{ps}	9	1.6 (250)	2.8 (250)	0.8 (250)	12.4 (250)	8.8 (250)	11.2 (250)
$MS1_{ps}$	2	**58.0** (250)	21.2 (250)	**69.6** (250)	44.8 (250)	12.4 (250)	**57.2** (250)
$MS1_{ps}$	3	**82.0** (250)	24.4 (250)	**89.2** (250)	**73.2** (250)	13.6 (250)	**83.2** (250)
$MS1_{ps}$	4	**91.2** (250)	29.2 (250)	**94.0** (250)	**85.6** (250)	15.6 (250)	**89.2** (250)
$MS1_{ps}$	5	**93.2** (250)	29.2 (250)	**98** (250)	**88.0** (250)	20.4 (250)	**95.6** (250)
$MS1_{ps}$	6	**95.6** (250)	32.8 (250)	**99.2** (250)	**91.6** (250)	21.6 (250)	**98** (250)
$MS1_{ps}$	7	**98.0** (250)	34.8 (250)	**99.6** (250)	**95.6** (250)	26.4 (250)	**98.8** (250)
$MS1_{ps}$	8	**99.6** (250)	35.6 (250)	**100.0** (250)	**98.4** (250)	24.4 (250)	**99.2** (250)
$MS1_{ps}$	9	**100.0** (250)	21.6 (250)	**100.0** (250)	**100.0** (250)	12.4 (250)	**100.0** (250)
$MS2_{ps}$	2	**74.8** (250)	26.4 (250)	**75.6** (250)	**63.2** (250)	18.0 (250)	**68.0** (250)
$MS2_{ps}$	3	**90.4** (250)	39.6 (250)	**94.4** (250)	**80.8** (250)	24.8 (250)	**87.2** (250)
$MS2_{ps}$	4	**93.6** (250)	41.2 (250)	**98.4** (250)	**89.6** (250)	28.0 (250)	**96.4** (250)
$MS2_{ps}$	5	**99.2** (250)	47.2 (250)	**99.2** (250)	**97.2** (250)	37.6 (250)	**98.4** (250)
$MS2_{ps}$	6	**98.8** (250)	**55.6** (250)	**100.0** (250)	**97.6** (250)	41.6 (250)	**100.00** (250)
$MS2_{ps}$	7	**99.2** (250)	**54.0** (250)	**99.6** (250)	**98.4** (250)	40.8 (250)	**99.6** (250)
$MS2_{ps}$	8	**99.6** (250)	**51.6** (250)	**100.0** (250)	**98.8** (250)	38.4 (250)	**100.0** (250)
$MS2_{ps}$	9	**100.0** (250)	47.2 (250)	**100.0** (250)	**99.6** (250)	35.6 (250)	**100.0** (250)
M_{pL}	2	2.8 (250)	4.4 (250)	4.4 (250)	7.6 (250)	5.6 (250)	7.2 (250)
M_{pL}	3	4.4 (250)	4.4 (250)	3.6 (250)	7.2 (250)	6.8 (250)	4.8 (250)
M_{pL}	4	7.2 (250)	4.4 (250)	6.4 (250)	5.6 (250)	6.0 (250)	10.0 (250)
M_{pL}	5	3.6 (250)	4.0 (250)	6.8 (250)	4.0 (250)	6.0 (250)	6.4 (250)
M_{pL}	6	5.2 (250)	4.4 (250)	5.2 (250)	4.8 (250)	7.6 (250)	6.8 (250)
M_{pL}	7	3.6 (250)	2.8 (250)	4.8 (250)	3.2 (250)	5.2 (250)	5.6 (250)
M_{pL}	8	3.2 (250)	4.8 (250)	4.0 (250)	5.6 (250)	3.2 (250)	5.2 (250)
M_{pL}	9	6.0 (250)	3.6 (250)	6.0 (250)	6.4 (250)	4.0 (250)	4.0 (250)
S_{pL}	2	2.0 (250)	4.4 (250)	2.4 (250)	32.0 (250)	21.2 (250)	24.4 (250)
S_{pL}	3	1.6 (250)	5.6 (250)	2.4 (250)	19.6 (250)	12.4 (250)	15.2 (250)
S_{pL}	4	2.8 (250)	5.2 (250)	2.0 (250)	12.0 (250)	10.0 (250)	10.8 (250)
S_{pL}	5	2.8 (250)	2.8 (250)	1.2 (250)	9.6 (250)	7.6 (250)	8.4 (250)
S_{pL}	6	3.2 (250)	4.8 (250)	3.6 (250)	13.2 (250)	9.6 (250)	8.0 (250)
S_{pL}	7	4.0 (250)	2.4 (250)	3.2 (250)	14.0 (250)	6.4 (250)	11.2 (250)
S_{pL}	8	1.6 (250)	2.8 (250)	4.0 (250)	13.2 (250)	4.4 (250)	12.0 (250)
S_{pL}	9	2.0 (250)	2.4 (250)	2.4 (250)	12.4 (250)	12.0 (250)	12.8 (250)
$MS1_{pL}$	2	44.0 (250)	16.4 (250)	42.0 (250)	34.0 (250)	13.2 (250)	31.6 (250)
$MS1_{pL}$	3	54.0 (250)	24.0 (250)	**56.0** (250)	44.0 (250)	15.6 (250)	43.2 (250)
$MS1_{pL}$	4	**69.2** (250)	23.6 (250)	**66** (250)	**58.8** (250)	14.8 (250)	**59.6** (250)
$MS1_{pL}$	5	**74.4** (250)	25.2 (250)	**75.6** (250)	**65.6** (250)	15.6 (250)	**63.6** (250)
$MS1_{pL}$	6	**77.2** (250)	29.6 (250)	**78.8** (250)	**71.6** (250)	17.2 (250)	**69.2** (250)
$MS1_{pL}$	7	**84.8** (250)	35.2 (250)	**86.8** (250)	**74.8** (250)	26.4 (250)	**77.2** (250)
$MS1_{pL}$	8	**89.2** (250)	32.8 (250)	**87.2** (250)	**82** (250)	22.4 (250)	**78.8** (250)
$MS1_{pL}$	9	**82.8** (250)	32.8 (250)	**85.2** (250)	**73.2** (250)	22 (250)	**74.4** (250)
$MS2_{pL}$	2	47.6 (250)	32.8 (250)	48.4 (250)	33.2 (250)	19.6 (250)	37.2 (250)
$MS2_{pL}$	3	**70.8** (250)	27.2 (250)	**64.0** (250)	**56** (250)	17.2 (250)	48.8 (250)
$MS2_{pL}$	4	**76.8** (250)	41.2 (250)	**74** (250)	**66.0** (250)	31.2 (250)	**63.2** (250)
$MS2_{pL}$	5	**81.6** (250)	39.2 (250)	**80.8** (250)	**73.2** (250)	27.2 (250)	**70.0** (250)
$MS2_{pL}$	6	**89.6** (250)	44.4 (250)	**86.4** (250)	**81.2** (250)	34.0 (250)	**80.8** (250)
$MS2_{pL}$	7	**94.8** (250)	**52.4** (250)	**89.6** (250)	**86.0** (250)	37.6 (250)	**85.6** (250)
$MS2_{pL}$	8	**94.8** (250)	49.6 (250)	**92.8** (250)	**87.2** (250)	33.2 (250)	**87.6** (250)
$MS2_{pL}$	9	**90.4** (250)	**58.4** (250)	**92.8** (250)	**85.2** (250)	46.0 (250)	**87.2** (250)

3.5 Adapted tests to distinguish a mover-stayer structure from memory

3.5.1 Test WBWA adapted for memory

We first attempted to adapt Test WBWA so that it would not react to a mover-stayer structure. The usual WBWA(i,r) contingency tables by state and occasion are modified by removing the animals who are in the same state at the previous and current occasions or at the current and future occasions since they could be potential stayers. In other words, the row and column corresponding to the current state are deleted from the original WBWA (i,r) contingency table (recall Figure 3.2). Consequently, this adapted test can only be used for a capture-recapture experiment with at least 3 live states. The results of the adapted test are shown in Table 3.19 for the global test (obtained as usual, by summing up the chi-square statistics resulting from the adapted tests by state and occasion, which are not presented here since they do not provide additional information). The adapted WBWA is no longer sensitive to a mover-stayer structure (around 5% of significant results), whilst it retains its high power to detect memory (100% of significant results for the considered scenarios). However it still lacks specificity since it remains sensitive to heterogeneity in preferences (64.8% for P2G, 72% for A2G) and heterogeneity in movement (100% of significant results).

3.5.2 Test of positive association adaptations

Not taking into account potential memory

We modified the test of positive association so that it would target a mover-stayer structure more specifically and not react to memory: our first solution is to not take into account the occasions before and after the middle occasion if the animal is in the same state at these 2 occasions, since this could potentially

Table 3.19: Global Test WBWA adapted for memory, high percentage of significant results in bold (> 50%)

Scenario	% (N)
M	6.0 (250)
S	0.0 (250)
MO	4.8 (250)
P	0.0 (250)
A	3.6 (250)
SD1	4.8 (250)
SD2	3.2 (250)
SD3	0.0 (250)
Mem1	**100.0** (250)
Mem2	**100.0** (250)
MS1	5.2 (250)
MS2	4.8 (250)
P2G	**64.8** (250)
A2G	**72.0** (250)
HM	**100.0** (250)
M_ps	2.4 (250)
S_ps	0.0 (250)
$MS1_ps$	3.6 (250)
$Ms2_ps$	0.4 (250)
M_pL	6.4 (250)
S_pL	0.0 (250)
$MS1_pL$	3.6 (250)
$MS2_pL$	1.2 (250)

be reflective of memory. This movement is also removed from the number of possible movements. This modification is illustrated in Table 3.20 for a toy example. Note that this test was performed on animals captured at least 5 times in order to avoid situations with 0 possible movements. The results of this adaptation are shown in Tables 3.21 for the summarised test. This first adaptation of the test of positive association does not seem successful since the test is still sensitive to memory (78.4% of significant results for Mem1). We note however, that the adapted test is sensitive to none of scenarios considered apart from the mover-stayer and memory scenarios, which is an improvement compared to the non adapted test.

Not taking into account memory, but keeping potential stayer information

Our second proposed solution is to not take into account the occasions before and after the middle occasion if the animal is in the same state at these 2 occasions, when the state at the middle occasion is different (potential memory). However, these occasions are retained if the animal is at the same state at the three occasions since this is potentially reflective of stayers. Again, we use only animals captured at least 5 times in order to make sure that we have at least one informative previous and future movement. A toy example of this second adaptation is given in Table 3.22.

Table 3.20: Toy example, modified positive association test, version 1, NM denotes the number of movements and Max the maximum number of possible movements.

ID	Non-zero capture history	Previous movements		Future movements	
		NM	Max	NM	Max
1	3 2 **3** 3 1	2	2	1	2
3	1 1 **1** 1 3 1	0	1	2	2
11	2 1 **2** 1 1 1	1	1	0	2

Table 3.21: Summarised test of positive association,adapted, conservative, at least 2 informative movements (animals captured at least 5 times), version 1, high percentage of significant results in bold (> 50%)

Scenario	% (N)
M	0.0 (250)
S	0.0 (250)
MO	0.0 (250)
P	0.0 (250)
A	0.0 (250)
SD1	0.0 (250)
SD3	0.0 (250)
Mem1	**78.4** (250)
Mem2	4.0 (250)
MS1	**100.0** (250)
MS2	**100.0** (250)
P2G	0.0 (250)
A2G	0.0 (250)
HM	0.0 (250)
M_{ps}	0.0 (250)
S_{ps}	0.0 (250)
$MS1_{ps}$	**100.0** (250)
$MS2_{ps}$	**100.0** (250)
M_{pL}	0.0 (250)
S_{pL}	0.0 (250)
$MS1_{pL}$	**90.0** (250)
$MS2_{pL}$	**95.2** (250)
M_t	0.0 (250)
S_t	0.0 (250)
$MS1_t$	**100.0** (250)
$MS2_t$	**100.0** (250)
$SD1_t$	0.0 (250)

Table 3.22: Toy example, modified positive association test, version 2, NM denotes the number of movements and Max the maximum number of possible movements.

ID	Non-zero capture history	Previous movements		Future movements	
		NM	Max	NM	Max
1	3 2 **3** 1 3	2	2	2	2
3	1 1 **1** 1 3 1	0	2	2	3
11	2 1 **2** 1 1 1	1	1	0	2

The results obtained with this strategy are given, for the summarised test, in Table 3.23. Like the first proposed adaptation, this second adaptation is very powerful at detecting mover-stayer structures (100% of significant results in all considered situations) and very sensitive to memory (around 100% too). Hence, it does not allow a mover-stayer structure to be differentiated from memory. One again, the adapted test is not sensitive at all to all other scenarios considered.

3.5.3 The solution to detecting a mover-stayer structure: using two adapted tests in conjunction

Based on the results from Sections 3.5.1 and 3.5.2, a possible solution for detecting a mover-stayer structure would be to combine the adapted WBWA with the second adaptation of the test of positive association. Indeed, the adapted WBWA is not affected by a mover-stayer scenario, and reacts strongly to heterogeneity in movement and preferences as well as memory; whilst the second adaptation of the positive association test is sensitive only to memory and a mover-stayer structure (note that we chose the second adaptation rather than the first because it was slightly more powerful). The possible outcomes are shown in Table 3.24. Both tests used together facilitate the detection of a mover-stayer structure, and the presence of memory, separately from other phenomena such as heterogeneous groups of preference or movement among

Table 3.23: Summarised test of positive association, conservative, adapted, version 2, high percentage of significant results in bold (> 50%)

Scenario	% (N)
M	0.4 (250)
S	0.0 (250)
MO	0.0 (250)
P	0.4 (250)
A	0.0 (250)
SD1	0.0 (250)
SD3	0.0 (250)
Mem1	**100.0** (250)
Mem2	**91.2** (250)
MS1	**100.0** (250)
MS2	**100.0** (250)
P2G	1.2 (250)
A2G	0.0 (250)
HM	0.0 (250)
M_ps	0.0 (250)
S_ps	0.0 (250)
$MS1_ps$	**100.0** (250)
$MS2_ps$	**100.0** (250)
M_pL	0.4 (250)
S_pL	0.4 (250)
$MS1_pL$	**100.0** (250)
$MS2_pL$	**100.0** (250)
M_t	0.0 (250)
S_t	0.0 (250)
$MS1_t$	**100.0** (250)
$MS2_t$	**100.0** (250)
$SD1_t$	0.0 (250)

the animals. Indeed, if both tests yield significant results, this is indicative of memory. A significant result for the adapted WBWA alone is indicative of heterogeneity in movement or preferences; whilst a significant result for only the adapted test of positive association is indicative of the existence of a mover-stayer structure.

Table 3.24: Test results' significance and possible conclusions

Adapted WBWA (Y/N)	Adapted positive association, version 2 (Y/N)	Conclusion
Y	N	Heterogeneity in movement/preferences
Y	Y	Memory
N	Y	Mover-stayer structure

3.6 Application: Canada geese

The famous Canada geese dataset from Hestbeck et al. (1991) is very often used as an illustration of memory (see for example Pradel et al., 2005; Rouan et al., 2009); it consists of 21,435 migrant geese individually marked with neck-bands and re-observed at their wintering locations each year, between 1984 and 1989 (Hestbeck et al., 1991; Rouan et al., 2009). These wintering sites constituted the states in the capture-recapture experiment: 1 denoted mid-Atlantic (New York, Pennsylvania, New Jersey), 2 Chesapeake (Delaware, Maryland, Virginia), and 3 Carolinas (North and South Carolina). Due to the new findings regarding the conclusions drawn from Test WBWA, we re-examine the geese dataset, using the combination of adapted tests to determine whether we still reach the same conclusion of memory.

Table 3.25 shows that the adapted test of positive association yields a significant result ($p = 0.01$); we have also detailed the test split by state, mainly to show how many animals were used for the adapted test, which is based on animals captured at least 5 times and we note that the number

of animals effectively used for the test is quite low compared to the size of the original dataset. The adapted WBWA test also yields a significant result ($p < 0.001$). According to Table 3.24, there is significant evidence that the geese display memory, which confirms the previous findings.

As a simple verification, we fitted a few simple models with different settings of survival and capture, for both memory and a mover-stayer structure, in order to check whether, for equivalent parametrisation of survival and capture probabilities, the model with memory was selected as a better model than the mover-stayer model. Note that we did not go through an exhaustive model fitting process since we aimed only to compare a memory model and a mover-stayer model fitted to the geese dataset. The models were fitted using program E-SURGE (Multi**E**vent **SUR**vival **G**eneralized **E**stimation) (Choquet et al., 2009b).

Both the memory model and the mover-stayer model are not multi-state models. Rather, they are both more general multievent models, which we briefly touched upon in Chapter 1 (Section 1.2). Multievent models will be presented in detail in Chapter 4; however we give here a brief overview of the tools necessary to comprehend the model fitting performed in this section.

Multievent models are more general than multi-state models in that they allow uncertainty in the state assignment: the observations upon capture constitute events while the states are underlying. For example, being a mover or a stayer is not a characteristic observable upon capture of the animal, and the observation "seen in colony 3" will be modelled as possibly resulting from an animal in states "mover in colony 3" or "stayer in colony 3". Multievent models are conditioned on first capture (McCrea and Morgan, 2014, p.100) and defined by the following parameters (again, further explanation will follow in Chapter 4): initial state probabilities, survival and transition probabilities, and event probabilities. Program E-SURGE is based on a general multi-event formulation of the model structured by the user and constrained using a language

Table 3.25: Canada geese: adapted test of positive association(version 2) for a mover-stayer structure, by state and summarised; $\hat{\gamma}$ denotes the gamma estimate, z(C) and pval (C) respectively denote the test-statistic and p-value for the adapted test of positive association with C reminding that this test is conservative, n denotes the number of animals used for the test, S in the state column indicates the summarised test.

$\hat{\gamma}$	z(C)	pval (C)	n	state
0.36	0.85	0.20	66	1
0.72	2.22	**0.01**	81	2
NA	NA	NA	21	3
-	2.17	**0.01**	-	S

GEMACO (see Choquet and Nogue, 2006, for details).

Various models have been proposed to account for memory, we chose to fit the Pradel memory model (Pradel, 2005), where the initial probabilities of the animals are dependent on their previous (unknown) site (Rouan et al., 2009), so that all animals follow the same survival-transition matrix. To fit the memory models in E-SURGE, we followed the step-by-step tutorial given in the web-appendix from Rouan et al. (2009), available at `https://static-content.springer.com/esm/art%3A10.1198%2Fjabes.2009.06108/MediaObjects/13253_2009_140300338_MOESM1_ESM.pdf`, only separating the steps of survival and transitions, as per Choquet et al. (2009b). Recall that the structure of the transition matrix for a memory model was given in Section 3.4.1, the initial states probability vector $\boldsymbol{\Pi_t}$, the survival matrix $\boldsymbol{\Phi_t}$ and the event matrix $\boldsymbol{B_t}$ are given below. For notation purposes, we follow the convention used in E-SURGE of denoting by * the probabilities equal to one minus the row-sum of the remaining terms. The labels of the columns and rows are denoted explicitly for clarity: for $\boldsymbol{\Pi_t}$ the columns represent the initial states, for $\boldsymbol{\Phi_t}$ the rows and columns represent the pairs of previous and current sites and finally for $\boldsymbol{B_t}$ the rows represent the underlying state of the animal whereas the columns represent the observations or in other words, the data collected

from a capture-recapture experiment.

$$\mathbf{\Pi_t} = \begin{array}{c} \begin{matrix} \scriptstyle(1,1) & \scriptstyle(1,2) & \scriptstyle(1,3) & \scriptstyle(2,1) & \scriptstyle(2,2) & \scriptstyle(2,3) & \scriptstyle(3,1) & \scriptstyle(3,2) & \scriptstyle(3,3) & \scriptstyle\dagger \end{matrix} \\ \begin{bmatrix} \pi_t^{11} & \pi_t^{12} & \pi_t^{13} & \pi_t^{21} & \pi_t^{22} & \pi_t^{23} & \pi_t^{31} & \pi_t^{32} & * & 0 \end{bmatrix} \end{array}.$$

$$\mathbf{\Phi_t} = \begin{array}{c} \\ \scriptstyle(1,1) \\ \scriptstyle(1,2) \\ \scriptstyle(1,3) \\ \scriptstyle(2,1) \\ \scriptstyle(2,2) \\ \scriptstyle(2,3) \\ \scriptstyle(3,1) \\ \scriptstyle(3,2) \\ \scriptstyle(3,3) \\ \scriptstyle\dagger \end{array} \begin{array}{c} \begin{matrix} \scriptstyle(1,1) & \scriptstyle(1,2) & \scriptstyle(1,3) & \scriptstyle(2,1) & \scriptstyle(2,2) & \scriptstyle(2,3) & \scriptstyle(3,1) & \scriptstyle(3,2) & \scriptstyle(3,3) & \scriptstyle\dagger \end{matrix} \\ \begin{bmatrix} \phi_t^{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * \\ 0 & \phi_t^{12} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * \\ 0 & 0 & \phi_t^{13} & 0 & 0 & 0 & 0 & 0 & 0 & * \\ 0 & 0 & 0 & \phi_t^{21} & 0 & 0 & 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 & \phi_t^{22} & 0 & 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 0 & \phi_t^{23} & 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & \phi_t^{31} & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_t^{32} & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_t^{33} & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}.$$

$$\boldsymbol{B_t} = \begin{array}{c} \\ \scriptstyle(1,1) \\ \scriptstyle(1,2) \\ \scriptstyle(1,3) \\ \scriptstyle(2,1) \\ \scriptstyle(2,2) \\ \scriptstyle(2,3) \\ \scriptstyle(3,1) \\ \scriptstyle(3,2) \\ \scriptstyle(3,3) \\ \scriptstyle\dagger \end{array} \begin{array}{c} \begin{matrix} \scriptstyle 0 & \scriptstyle 1 & \scriptstyle 2 & \scriptstyle 3 \end{matrix} \\ \begin{bmatrix} * & p_t^1 & 0 & 0 \\ * & 0 & p_t^2 & 0 \\ * & 0 & 0 & p_t^3 \\ * & p_t^1 & 0 & 0 \\ * & 0 & p_t^2 & 0 \\ * & 0 & 0 & p_t^3 \\ * & p_t^1 & 0 & 0 \\ * & 0 & p_t^2 & 0 \\ * & 0 & 0 & p_t^3 \\ * & 0 & 0 & 0 \end{bmatrix} \end{array}.$$

For the mover-stayer model, we used a mixture model with two groups of animals characterised by different transition structures. Movement between groups is not allowed since animals are assumed to be intrinsically either movers or stayers. The six live states resulting from this definition are: Stayer in location 1, 2, or 3 and Mover in 1, 2, or 3, respectively denoted by S1, S2,

S3, M1, M2, and M3. The characteristic matrices of the mover-stayer model are given below.

$$\mathbf{\Pi_t} = \begin{array}{c} \\ \end{array} \begin{array}{c} \begin{array}{ccccccc} (S1) & (S2) & (S3) & (M1) & (M2) & (M3) & \dagger \end{array} \\ \begin{bmatrix} \pi_t^{S1} & \pi_t^{S2} & \pi_t^{S3} & \pi_t^{M1} & \pi_t^{M2} & * & 0 \end{bmatrix} \end{array}.$$

$$\mathbf{\Phi_t} = \begin{array}{c} \\ (S1) \\ (S2) \\ (S3) \\ (M1) \\ (M2) \\ (M3) \\ \dagger \end{array} \begin{array}{c} \begin{array}{ccccccc} (S1) & (S2) & (S3) & (M1) & (M2) & (M3) & \dagger \end{array} \\ \begin{bmatrix} \phi_t^{S1} & 0 & 0 & 0 & 0 & 0 & * \\ 0 & \phi_t^{S2} & 0 & 0 & 0 & 0 & * \\ 0 & 0 & \phi_t^{S3} & 0 & 0 & 0 & * \\ 0 & 0 & 0 & \phi_t^{M1} & 0 & 0 & * \\ 0 & 0 & 0 & 0 & \phi_t^{M2} & 0 & * \\ 0 & 0 & 0 & 0 & 0 & \phi_t^{M3} & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}.$$

$$\mathbf{\Psi_t} = \begin{array}{c} \\ (S1) \\ (S2) \\ (S3) \\ (M1) \\ (M2) \\ (M3) \\ \dagger \end{array} \begin{array}{c} \begin{array}{ccccccc} (S1) & (S2) & (S3) & (M1) & (M2) & (M3) & \dagger \end{array} \\ \begin{bmatrix} \psi_t^{S1,1} & \psi_t^{S1,2} & * & 0 & 0 & 0 & 0 \\ \psi_t^{S2,1} & \psi_t^{S2,2} & * & 0 & 0 & 0 & 0 \\ \psi_t^{S3,1} & * & \psi_t^{S3,3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \psi_t^{M1,1} & \psi_t^{M1,2} & * & 0 \\ 0 & 0 & 0 & \psi_t^{M2,1} & \psi_t^{M2,2} & * & 0 \\ 0 & 0 & 0 & \psi_t^{M3,1} & * & \psi_t^{M3,3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}.$$

$$\boldsymbol{B_t} = \begin{array}{c} \\ (S1) \\ (S2) \\ (S3) \\ (M1) \\ (M2 \\ (M3) \\ \dagger \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} * & p_t^1 & 0 & 0 \\ * & 0 & p_t^2 & 0 \\ * & 0 & 0 & p_t^3 \\ * & p_t^1 & 0 & 0 \\ * & 0 & p_t^2 & 0 \\ * & 0 & 0 & p_t^3 \\ * & 0 & 0 & 0 \end{bmatrix} \end{array}.$$

It should be noted that a model with two groups of animals characterised by different transition matrices is actually appropriate for a broader spectrum of models than just mover-stayer, they can be used for other situations of heterogeneity in movement or preferences. However, fitting a strictly mover-stayer model would require additional constraints of the form $\{\psi^{S1,1} > \psi^{M1,1}$ and $\psi^{S2,2} > \psi^{M2,2}$ and $\psi^{S3,3} > \psi^{M3,3}\}$ that are not necessarily straightforward to implement in pre-existing software such as E-SURGE. Thus it is less likely to be routinely fitted in practice than the more general mixture model.

We are aware that fitting memory models involves numerous identifiability issues (Rouan et al., 2009; Cole et al., 2014), this issue is out of the scope of this thesis and we did not dwell on it. Note however, that E-SURGE uses a built-in tool which provides a numerical estimate of the number of estimable parameters in the model (Choquet and Nogue, 2006).

The models fitted are presented in Table 3.26, they are defined following the GEMACO terminology used in program E-SURGE (Choquet et al., 2009b): c indicates that the probabilities are constant over time and states, to indicates that the probabilities are different along the columns, $from$ indicates that the probabilities are different along the rows, $from.to$ indicates that they differ along both rows and columns (used to constrain the transition probabilities to be dependent on both the state of departure and the state of arrival). The initial probabilities are conventionally fitted as both state and time-dependent: $to.t$. The best model was chosen using Akaike's information criterion, the AIC, defined as $-2\log\{L(\hat{\theta})\} + 2d$, $log\{L(\hat{\theta})\}$ denotes the log-likelihood evaluated at its maximum and d denotes the number of estimable parameters (Burnham and Anderson, 2002, p.61). We were concerned about using the AIC because the memory models and the mover-stayer models are both based on a different number of underlying states, which also have different meanings. However these models are all fitted to the same representation

of the dataset (no conditional likelihood or different grouping), therefore the AIC can be used for model comparison(Burnham and Anderson, 2002, p.81). For reassurance, we fitted both a memory model and a mover-stayer model to a randomly chosen simulated dataset from scenario MS2 (with all probabilities constant except for the state-dependent transition probabilities). The best model chosen by the AIC was the mover-stayer model (AIC=26979.19, versus AIC=27318.27 for the memory model).

For the goose dataset, the memory model was consistently found to be better than the mixture model, for equivalent types of parameter dependencies, which is in agreement with the results of the tests presented in Table 3.25.

3.7 Discussion

To summarise, this chapter extended the scope of existing Test WBWA, which is currently used as a test for memory. We showed that the test is actually sensitive to other violations of the homogeneity in transition assumption, such as heterogeneity in movement patterns, or in preference/avoidance, and a mover-stayer structure. Thus, when Test WBWA produces a significant result, these alternative models should be considered and fitted if they are biologically sensible in addition to the memory model.

We also examined the properties of a positive association test to detect a mover-stayer structure, directly extending the test for heterogeneity in capture

Table 3.26: Model fitting: Memory and mixture, Canada Geese dataset

Model	N parameters	Deviance	AIC	Delta(AIC)
Memory: $\pi(to.t), p(to), \phi(from), \psi(from.to)$	72	115904.47	116048.47	0.00
Memory: $\pi(to.t), p(c), \phi(from), \psi(from.to)$	70	115968.92	116108.92	60.45
Mixture: $\pi(to.t), p(to), \phi(from), \psi(from.to)$	48	116023.92	116119.92	71.45
Mixture: $\pi(to.t), p(c), \phi(from), \psi(from.to)$	46	116053.91	116145.91	97.44
Memory: $\pi(to.t), p(c), \phi(c), \psi(from.to)$	62	116077.41	116201.41	152.93
Mixture: $\pi(to.t), p(c), \phi(c), \psi(from.to)$	41	116135.08	116217.08	168.61

derived in Chapter 2. However, this test lacked specificity, and was very sensitive to both memory and mover-stayer structure, whilst also being slightly affected by some of the other scenarios.

Consequently, we modified the current Test WBWA to derive a test that was no longer sensitive to a mover-stayer structure. We also adapted the test of positive association, so that it became sensitive only to memory and a mover-stayer structure. Finally we used both the adapted Test WBWA and the adapted test of positive association to detect and distinguish between memory, mover-stayer structure and heterogeneity in movement or preferences. This combined tool can be used for a capture-recapture experiment with at least 3 live states and 5 capture occasions. The level of dependence between the adapted WBWA test and the adapted test of positive association is unclear at this stage. To be cautious, we would recommend using a correction for multiple testing, such as the Bonferroni correction which consists of dividing the chosen level by the number of hypotheses tested (see for example Sokal and Rohlf, 2012, p.239).

The main advantage of the new tool is that it provides more specific information, without needing any model fitting and is very powerful in good conditions. Its main limitation at the current stage is the requirement for large sample sizes so that there is enough testable data, particularly for the test of positive association. Further investigations are needed in order to adapt this tool for smaller sample sizes. For instance, in the same way as for Chapter 2, a permutation test could be explored.

Finally, we keep in mind that biological behaviours are, by essence, more complex than simulated scenarios involving only a clear-cut phenomenon. For instance, animals could present long-term memory; they could also change their moving behaviour over time, if for example, they scout until they find a nice colony to settle in, they would first be movers and then stayers. It would be of interest to research how the existing tests and the new tool react in these

types of situation.

Chapter 4

Testing the underlying state structure for partially observed multi-state data

4.1 Introduction

The states defined in a multi-state model are observable and assigned with certainty (Kendall, 2004). Partial observations occur when the state cannot be determined for a proportion of animals at a given sampling occasion (Conn and Cooch, 2009). In monomorphic species, male and female individuals do not present obvious physical differences. As a result, individuals need to be directly handled or genetically tested to determine their gender. This is not always possible and then, the sex needs to be inferred based solely on the animal's behaviour (Genovart et al., 2012). If, at a given occasion, there are individuals for which the sex cannot be ascertained and is left as "unknown", these constitute partial observations. For other individuals, sex may be assigned with uncertainty. In animal epidemiology, where health status constitutes the states, such as the study of avian malaria (Lachish et al., 2011)

or conjunctivitis in *Carpodacus mexicanus* Müller house finches (Conn and Cooch, 2009), some animals' health status cannot be determined when direct testing via blood samples, for instance, is not possible.

As mentioned previously in Section 3.6 from Chapter 3, multievent models were developed by Pradel (2005) in order to take into account state uncertainty: events are observed whilst the states are underlying. Multievent models are Hidden Markov Models, they are layered: a hidden Markov process, generally assumed of order 1, governs the movement between the underlying states and the events are generated by these states. Note that the set of underlying states, although not directly observable, is defined by the user according to their expertise and the question of interest. Multievent models condition on first capture and form a general framework which allow the relaxation of standard multi-state or CJS model assumptions. Example situations that can be modelled by multievent models include transience, heterogeneity or memory.

The likelihood of a multievent model is given in Equation 4.1, using the following notation:

- N the number of observed individuals,
- R the number of states defined by the user (including the state "dead"),
- E the number of events (including the event "not captured"),
- $\mathbf{D}(\boldsymbol{\theta})$ the diagonal matrix with diagonal elements equal to the elements of $\boldsymbol{\theta}$,
- f_k the first encounter occasion for animal k,
- T the number of occasions in the capture-recapture experiment,
- $\mathbf{1}_R$ a column vector of R ones,
- $\mathbf{\Pi_t}$ the $1 \times R$ vector of initial state probabilities, with the rth element

being the probability that an individual is in state r when initially encountered at occasion t,

- $\mathbf{\Phi_t}$ the $R \times R$ survival matrix, with the diagonal terms ϕ_t^r denoting the probability that an animal in state r at time t survives until $t+1$ for $r < R$. For R, the "dead" state, diagonal element $\phi_t^R = 1$ and the Rth column is formed of the terms $1 - \phi_t^r$ for each row $r < R$.

- $\mathbf{\Psi_t}$ the $R \times R$ transition matrix, with the (r,s)th element being $\psi_t^{r,s}$, the probability that an animal is in state s at time $t+1$, given it was in state r at t and that it is alive at $t+1$.

- $\mathbf{B_t}$ the $R \times E$ event matrix with the (r,e)th element being the probability of observing event e for an animal in state r at time t. The notation $\mathbf{B}_t(.,e_{k,t})$ refers to the column of $\mathbf{B_t}$ corresponding to the event observed at time t for animal k.

The matrices $\mathbf{\Pi_t}$, $\mathbf{\Phi_t}$, $\mathbf{\Psi_t}$ and $\mathbf{B_t}$ are all row-stochastic and recall that they were illustrated for various examples in Chapter 3, Section 3.6.

$$L = \prod_{k=1}^{N} \left(\mathbf{\Pi}_{f_k} \mathbf{D}(\mathbf{B}_{f_k}(.,e_{f_k})) \left[\prod_{t=f_k+1}^{T} \mathbf{\Phi}_{t-1} \mathbf{\Psi}_{t-1} \mathbf{D}(\mathbf{B}_t(.,e_{k,t})) \right] \mathbf{1}_R \right) \tag{4.1}$$

This chapter focusses on multi-state capture-recapture data with partial observations (i.e. captured/sighted with state unknown), assuming that when states are assigned to observed animals, this is done without error/uncertainty. These data can be modelled as a special case of multievent model, and this chapter proposes a new diagnostic tool to assess whether the partial observations are actually generated from the states directly observed in the capture-recapture experiment.

Consider a capture-recapture experiment designed to study avian malaria, with the underlying states being defined as "healthy" (H), "infected with malaria" (I) and "dead"(†); the events recorded are "not captured" (NC), "observed

as healthy" (H), "observed as infected with malaria" (I) and "observed with health status unknown" (U), which constitutes the partial observations. If an animal is observed as healthy or infected with malaria, then this is actually in the corresponding underlying state. If an animal's health status is unknown, it could actually be either healthy or infected with malaria. Finally, when an animal is not captured, it could be in any of the three underlying states. This is a situation where partial observations are generated from the observed states; and it is illustrated for a given occasion in Figure 4.1.

An example of alternative scenario is represented in Figure 4.2, for the same health status example. Only this time, the partial observations correspond to animals who are actually in State C which represents "infected with conjunctivitis". For clarity purposes, Figure 4.2 zooms in on the observations made, leaving out the event "not captured" (and by extension the state "dead").

The test developed in this chapter builds on the approach utilised by Test M in the multi-state framework (Pradel et al., 2003). Test M is presented in Section 4.2. In Section 4.3, we show that if partial observations are generated only by the directly observable states, then animals partially observed at time i are a mixture of the animals observed in any of the observable states at that occasion. Based on this mixture property, a new test is developed to assess whether partial observations are actually generated by the observable states. The properties of this test are assessed using simulation in Section 4.4. Finally, in Section 4.5 the test is applied to the Canada geese dataset used in Chapter 3 as well as a dataset of greater flamingoes (*Phoenicopterus roseus*), to explore its performance in real-life situations. Finally, we conclude and discuss the findings in Section 4.6.

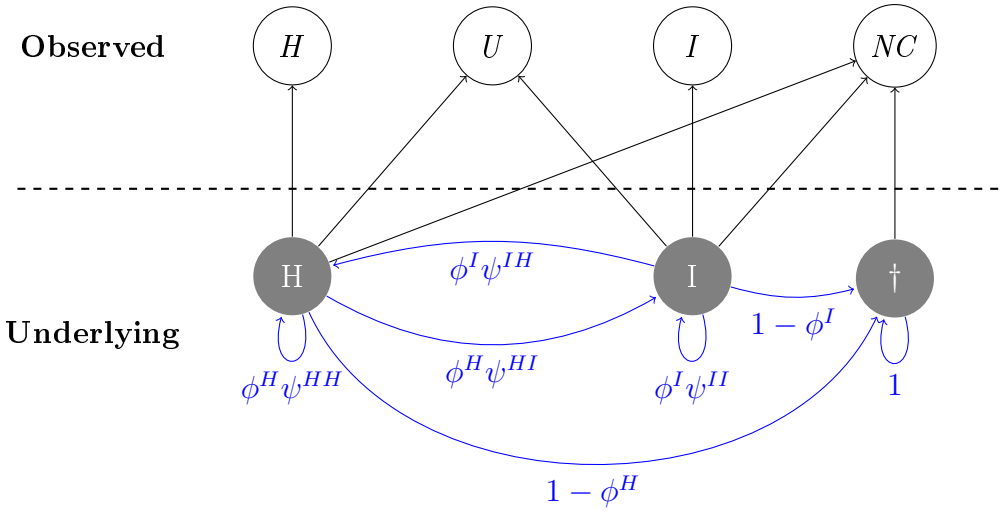


Figure 4.1: Underlying state structure illustration at a given occasion for a capture-recapture experiment with two live states H and I, directly observed without error and partial observations corresponding to animals that can be in either of these states; U denotes partial observations, ϕ^r denotes the survival probability in state r and ψ^{rs} the transition probabilities from state r to state s.

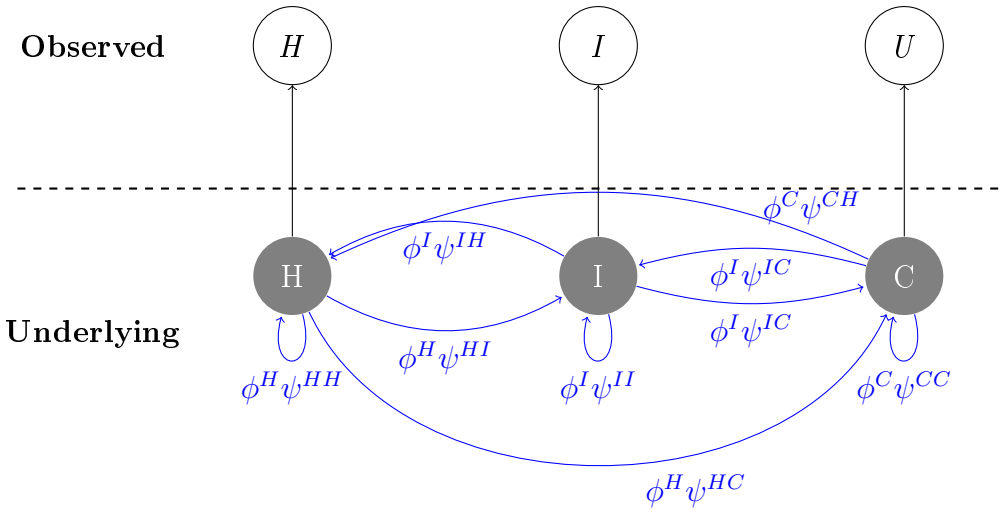


Figure 4.2: Illustration of the underlying state structure at a given occasion for a capture-recapture experiment with two live states H and I directly observed without error and partial observations corresponding to the additional state C which is never directly observable; U denotes partial observations, ϕ^r denotes the survival probability in state r and ψ^{rs} the transition probabilities from state r to state s.

4.2 Test M

In a multi-state framework, it is not possible to naturally extend Test 2, described in Chapter 2, which compares animals captured at time i to those not captured at time i. This is due to the fact that, in a multi-state setting, when an animal is not captured, its state is unknown. However, if the assumption of equal recapture probability at a given occasion for all animals in the same state is true, Pradel et al. (2003) noted that, amongst the animals still alive, those that are not captured at occasion i must be in one of the live states defined in the experiment. Thus, Pradel et al. (2003) demonstrated that the number of animals not captured at i, previously released and known to be alive after i, follows a conditional multinomial distribution, which is a finite mixture of the conditional multinomials followed by the number of animals seen at i in the different states and re-observed at least once. Pradel et al. (2003) also noted that although this mixture property is verified for the Arnason-Schwarz model, it is actually characteristic of the more general Jolly-Movement model. This is why the goodness-of-fit suite for the multi-state model assesses the fit of a Jolly-Movement model rather than an Arnason-Schwarz model, as mentioned in Chapter 3 (Section 3.2). Based on this mixture property, Pradel et al. (2003) used pooling strategies to derive a general mixture test at a given occasion i: Test M. At each occasion, Test M assesses whether animals previously released in a given state r, not captured at i and known to still be alive after i are consistent with being a mixture of the animals in the same conditions captured in either of the states at i. Table 4.1 shows the m-array terms constitutive of the table associated with Test M.

The probability density function of a finite mixture is defined as $g(x) = \sum_{c=1}^{C} \pi_c f_c(x)$ (see for example Everitt and Hand, 1981, p. 4); C is the number of mixture components, $f_c(x)$ the probability density function of component c, also termed basis distribution by Yantis et al. (1991), and π_c the

Table 4.1: Table of the m-array terms associated with Test M at occasion i, for a capture-recapture experiment with K occasions and R live states denoted by A to R. The mixtures M_1 to M_R, are each a mixture of the bases B_1 to B_R.

$\sum_{j=1}^{i-1} m_{j,i+1}^{AA}$	$\dots$	$\sum_{j=1}^{i-1} m_{j,i+1}^{AR}$	$\dots$	$\sum_{j=1}^{i-1} m_{j,K}^{AA}$	$\dots$	$\sum_{j=1}^{i-1} m_{j,K}^{AR}$	M_1
$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
$\sum_{j=1}^{i-1} m_{j,i+1}^{RA}$	$\dots$	$\sum_{j=1}^{i-1} m_{j,i+1}^{RR}$	$\dots$	$\sum_{j=1}^{i-1} m_{j,K}^{RA}$	$\dots$	$\sum_{j=1}^{i-1} m_{j,K}^{RR}$	M_R
$m_{i,i+1}^{AA}$	$\dots$	$m_{i,i+1}^{AR}$	$\dots$	$m_{i,K}^{AA}$	$\dots$	$m_{i,K}^{AR}$	B_1
$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
$m_{i,i+1}^{RA}$	$\dots$	$m_{i,i+1}^{RR}$	$\dots$	$m_{i,K}^{RA}$	$\dots$	$m_{i,K}^{RR}$	B_R

mixing probabilities associated with each component ($\sum_{c=1}^{C} \pi_c = 1$). Yantis et al. (1991) developed the Multinomial Maximum Likelihood Mixture approach (MMLM), which, as its name implies, is targeted to mixtures of multinomial distributions. The context of the MMLM approach is slightly different from the more common mixture model problems, in which only mixtures are sampled from and there is no direct information available from the underlying mixture components (Everitt and Hand, 1981, p. 2). In the MMLM setting, independent samples are available from both the mixtures and the components. Note that from here onwards, we will be using the terminology from Yantis et al. (1991) and Pradel et al. (2003), and terming these components *bases*.

The model structure corresponding to a MMLM setting is shown in Table 4.2 with the mixing probabilities denoted in blue and the bases cell-probabilities in red. The MMLM approach consists of two steps: estimating the bases cell probabilities and the mixing probabilities via maximum-likelihood. (Recall that each row represents a multinomial so the cell probabilities sum to 1 for each row.) The second step is assessing the goodness-of-fit of the hypothesised model structure (mixtures and bases) using a classical measure of comparison between observed and expected frequencies.

Test M is essentially an application of the MMLM approach to a capture-recapture setting, as illustrated by the m-array terms in Table 4.1, where the

Table 4.2: MMLM approach, Mixture and bases model structure associated with Table 4.1

$j=$	$i+1$			...	T			
$s=$	A	...	R	...	A	...	R	
	$\gamma_1 p_1^{B_1} + \ldots + \gamma_R p_1^{B_R}$	...	...	...	...	...	$\gamma_1 p_{R\times T}^{B_1} + \ldots + \gamma_R p_{R\times T}^{B_R}$	M_1
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$\pi_1 p_1^{B_1} + \ldots + \pi_R p_1^{B_R}$	...	...	...	...	...	$\pi_1 p_{R\times T}^{B_1} + \ldots + \pi_R p_{R\times T}^{B_R}$	M_R
	$p_1^{B_1}$	...	...	...	...	...	$p_{R\times T}^{B_1}$	B_1
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$p_1^{B_R}$	...	...	...	...	...	$p_{R\times T}^{B_R}$	B_R

mixtures are denoted by M and the bases by B. The associated model structure corresponds to the cell-probabilities given in Table 4.2 for the different mixtures and bases. Pradel et al. (2005) then partitioned Test M into two components: Test M.ITEC, which detects a short-term trap effect, by confronting the animals first re-observed at $i+1$ and those first re-observed at later times; and Test M.LTEC. In case of sparse data, the mixture rows may be pooled together, unlike the bases rows, which should never be pooled so that the hypothesised model structure is conserved. The table may also be pooled across the columns in any manner except for Test M.ITEC, where animals seen again at $i+1$ and those seen later should be kept separated (Choquet et al., 2005). Finally, note that Test M is not a test of independence so Fisher's exact test cannot be used in this case.

4.3 Testing the mixture property for partial observations

4.3.1 The mixture property of partial observations generated by only the observable states

If partial observations stem from the states directly observed in the experiment, then the set of underlying states is formed only of the observable states plus the dead state. In this section, we show that, in this case, the number of animals *partially observed* at i and seen again in a known state afterwards, follows a multinomial distribution which is a mixture of the multinomial distributions followed by the animals released in a *known state* at time i and seen again in a known state afterwards. The multinomial cells correspond to the time and state of the first re-observation in a known state after i.

Consider a capture-recapture experiment with T sampling occasions and R live states. If individuals are assigned to state r upon capture, this is done without uncertainty (see Section 4.1) and the corresponding event is denoted by r: "observed in state r". However, when an individual's state cannot be determined, the corresponding event is denoted by U: "observed with state unknown" and the animal can be in any one of the underlying R states.

In the presence of such partial observations, there is no longer a perfect match between observation and state; so the multi-state m-array (introduced in Chapter 3) are no longer sufficient-statistics. However, King and McCrea (2014) developed sufficient statistics for the more complex framework of partial observations which we have presented in this chapter. Below, we introduce the sufficient statistics, model parameters, and associated likelihood components that are required to prove the mixture property of the partial observations. We also introduce some additional notation in order to simplify the expression of the conditional probabilities needed.

- Sufficient statistic terms
  - $n_{t_1,t_2}^{r,z_{(t_1+1):(t_2-1)},s}$ denotes the number of animals observed at time t_1 in known state r, next observed in known state s at t_2 with partial capture history $z_{(t_1+1):(t_2-1)}$ between these two time points.
  - $w_{t_1,t_2}^{U,z_{(t_1+1):(t_2-1)},s}$ denotes the number of animals observed for the first time at t_1 in an unknown state, re-observed for the first time in known state s at time t_2 with partial capture history $z_{(t_1+1):(t_2-1)}$ between these two time points.
  - $v_{t_1}^r$, the number of animals observed in known state r at t_1 and never seen again in a known state (i.e. never seen again or only ever re-observed in an unknown state).
  - $b_{t_1}^U$, the number of animals first observed in an unknown state at t_1 and never seen again in a known state.
- Model parameters
  - ϕ_t^r: probability an individual in state r at time t survives until $t+1$, for $t = 1, \ldots, (T-1)$.
  - p_t^r the probability of recapture at time t for an individual in state r, for $t = 2, \ldots, T$.
  - $\psi_t^{r,s}$: probability an individual is in state s at time $t+1$ given that it was in state r at time t and is alive at $t+1$, for $t = 1, \ldots, T-1$
  - α_t^r: probability an individual is assigned to known state r given it was recaptured at time t, for $t = 2, \ldots, T$. $\beta_t^r = 1 - \alpha_t^r$ is then defined as the probability an individual is assigned as unknown (U) at time t given the individual is recaptured, and in state r at this time, for $t = 2, \ldots, T$
  - π_t^r: probability an individual is in state r at time t, given it was first observed in U at t, for $t = 1, \ldots, T-1$.

- Likelihood components associated with the sufficient statistics

  - The component associated with $n_{t_1,(t_2+1)}^{r,z_{(t_1+1):t_2},s}$ is $O_{t_1,t_2}^{r,s,z_{(t_1+1):t_2}}$, defined as the probability an individual released in the known state r at time t_1 is not observed in a known state between capture times (t_1+1) and t_2, with partial encounter history $z_{(t_1+1):t_2}$, and is observed at (t_2+1), in known state s: for $t_1 \leq t_2$, $O_{t_1,t_2}^{r,s,z_{(t_1+1):t_2}} = Q_{t_1,t_2}^{r,s,z_{(t_1+1):t_2}} p_{t_2+1}^s \alpha_{t_2+1}^s$.
    $Q_{t_1,t_2}^{r,s,z_{(t_1+1):t_2}}$ denotes the probability an individual released in the known state r at time t_1 is not observed in a known state between capture times (t_1+1) and t_2 with partial encounter history , $z_{(t_1+1):t_2}$, and is in state s at time (t_2+1).
    $$Q_{t_1,t_2}^{r,s,z_{(t_1+1):t_2}} = \begin{cases} \phi_{t_1}^r \psi_{t_1}^{r,s} & \text{if } t_1 = t_2 \\ \phi_{t_1}^r \sum_{l=1}^R \psi_{t_1}^{r,l} Z_{t_1+1}(l, h(t_1+1)) Q_{t_1+1,t_2}^{l,s,z_{(t_1+2):t_2}} & \text{if } t_1 < t_2 \end{cases}$$
    $Z_t(r, h(t))$ denotes the probability that an individual who is unobserved or observed in a partial state at time t, has the encounter history $h(t)$ at that time point, given that they are in state r: $Z_t(r, h(t)) = (1-p_t^r)^{\mathbb{1}_{\{h(t)=0\}}} (p_t^r \beta_t^r)^{\mathbb{1}_{\{h(t)\neq 0\}}}$.
  - The component associated with $w_{t_1,(t_2+1)}^{U,z_{(t_1+1):t_2},s}$ is the probability an individual initially observed in U at time t_1 is partially observed or unobserved until t_2 with partial encounter history $z_{(t_1+1):t_2}$, and observed at time t_2+1 in known state s:
    $\gamma_{t_1,t_2}^{U,s,z_{(t_1+1):t_2}} = \sum_{v=1}^R \pi_{t_1}^v O_{t_1,t_2}^{v,s,z_{(t_1+1):t_2}}$

- Additional notation

  - $\lambda_{t_1}^r$, the probability that an animal released at t_1 in known state r, is re-observed in a known state.
  - $\lambda_{t_1}^U$, the probability that an animal released at t_1 in an unknown state, is re-observed in a known state.

We illustrate in Table 4.3 how example individual capture histories are translated into the sufficient statistic terms presented above.

Consider, for example, occasion $i = 2$, for a capture-recapture experiment with two live states A and B. The number of animals released in state A at occasion 1 first re-captured in a known state at the different occasions, and those never seen again in a known state, follow a multinomial distribution. The same goes for those released in state B at occasion 1, those first released in an unknown state at occasion 1 etc. All these multinomials are described in terms of sufficient statistics and their associated cell-probabilities in Table 4.4 for a study with 4 occasions. Each odd row presents the sufficient statistics following a multinomial distribution, whilst each even row presents the cell-probabilities associated with that multinomial. From Table 4.4, it is apparent that some cell-probabilities relating to the animals not seen in a known state at occasion 2 are linear combinations of the cell-probabilities of animals seen in a known state at that same occasion (indicated in red). The quantities constituting the coefficients of the linear combinations are denoted in blue. Based on this observation, it follows that the terms denoted in black should be peeled-off from the table in order to construct a structure with mixtures and bases. Note that due to space constraints, we have not expanded the cell-probabilities corresponding to the animals first released in U at occasion 1 (first row of w-terms) so as to express them as a function of the bases terms denoted in red. However, it is a straightforward observation to make since the cell-probabilities of these w-terms correspond to a mixture of those associated with the n-terms denoted in blue in the same column. Building on this, it is possible to construct a general table for occasion i, based on the conditional multinomials obtained once the relevant terms (denoted in black in Table 4.4) are peeled off. When the number of sampling occasions increases, capture histories are longer and there is a great number of possible intermediate capture histories, formed of combinations of 0s and Us, before

Table 4.3: Illustrating how example individual capture histories contribute to the sufficient statistic terms, for a capture-recapture experiment with two observable states A, B and five sampling occasions. Partial observations are denoted by U. The elements of capture history determining the indices within the statistics are denoted in bold.

Capture History					sufficient statistic
U	**A**	U	U	**B**	$w_{1,2}^{U,-,A}$,$n_{2,5}^{A,UU,B}$
A	U	U	U	**A**	$n_{1,5}^{A,UUU,A}$
A	U	0	U	0	v_1^A
U	U	U	U	**B**	$w_{1,5}^{U,UUU,B}$
0	0	**U**	0	**1**	$w_{3,5}^{U,0,1}$
0	**A**	**B**	U	U	$n_{2,3}^{A,-,B}$, v_3^B
0	**U**	0	U	U	b_2^U

the first certain observation appears. In order to lower the chances of a sparse table, we opted to build the multinomials based on the time and state of the first known reobserved state, thus pooling over all possible intermediate capture histories.

Table 4.5 gives a general expression of the sufficient statistic terms relating to these conditional multinomials, for a capture-recapture experiment with R observable states, partial observations denoted U, and T sampling occasions whilst Table 4.6 displays the associated cell-probabilities. At occasion i, the conditional multinomials are thus formed by the animals released at $i-1$ in a known state, partially observed at i and next seen again in a known state by the end of the capture-recapture experiment, the animals first observed in U at $i-1$ or i and later re-observed in a known state, as well as the animals released at i in a known state and seen again in a known state. Again, in Tables 4.5 and 4.6, each row corresponds to a multinomial.

We demonstrate below the mixture property corresponding to the partial

Table 4.4: Multinomial distributions relative to the animals released before or at $i = 2$: sufficient-statistic terms (odd rows) and associated cell-probabilities (even rows). The terms constitutive of mixtures are denoted in blue whilst those constituting bases are denoted in red. The terms in black will be peeled out due to conditioning

$\mathbf{n}_{1,2}^{A,-,A}$	$\mathbf{n}_{1,2}^{A,-,B}$	$\mathbf{n}_{1,3}^{A,0,A}$ … $\mathbf{n}_{1,4}^{A,0U,B}$	$\mathbf{n}_{1,3}^{A,U,A}$	$\mathbf{n}_{1,3}^{A,U,B}$	$\mathbf{n}_{1,4}^{A,U0,A}$	$\mathbf{n}_{1,4}^{A,U0,B}$	$\mathbf{n}_{1,4}^{A,UU,A}$	$\mathbf{n}_{1,4}^{A,UU,B}$	v_1^A
$O_{1,1}^{A,A,-}$	$O_{1,1}^{A,B,-}$	$O_{1,2}^{A,A,0}$ … $O_{1,3}^{A,B,0U}$	$O_{1,2}^{A,A,U}$	$O_{1,2}^{A,B,U}$	$O_{1,3}^{A,A,U0}$	$O_{1,3}^{A,B,U0}$	$O_{1,3}^{A,A,UU}$	$O_{1,3}^{A,B,UU}$	$1-\lambda_1^A$
			$=Q_{1,2}^{A,A,U}p_3^A\alpha_3^A$	$=Q_{1,2}^{A,B,U}p_3^B\alpha_3^B$	$=Q_{1,3}^{A,A,U0}p_4^A\alpha_4^A$	$=Q_{1,3}^{A,B,U0}p_4^B\alpha_4^B$	$=Q_{1,3}^{A,A,UU}p_4^A\alpha_4^A$	$=Q_{1,3}^{A,B,UU}p_4^B\alpha_4^B$	
			$=\phi_1^A\psi_1^{AA}Z_2(A,U)Q_{2,2}^{A,A,-}p_3^A\alpha_3^A$	$=\phi_1^A\psi_1^{AA}Z_2(A,U)Q_{2,2}^{A,B,-}p_3^B\alpha_3^B$	$=\phi_1^A\psi_1^{AA}Z_2(A,U)Q_{2,3}^{A,A,0}p_4^A\alpha_4^A$	$=\phi_1^A\psi_1^{AA}Z_2(A,U)Q_{2,3}^{A,B,0}p_4^B\alpha_4^B$	$=\phi_1^A\psi_1^{AA}Z_2(A,U)Q_{2,3}^{A,A,U}p_4^A\alpha_4^A$	$=\phi_1^A\psi_1^{AA}Z_2(A,U)\,Q_{2,3}^{A,B,U}p_4^B\alpha_4^B$	
			$+\phi_1^A\psi_1^{AB}Z_2(B,U)Q_{2,2}^{B,A,-}p_3^A\alpha_3^A$	$+\phi_1^A\psi_1^{AB}Z_2(B,U)Q_{2,2}^{B,B,-}p_3^B\alpha_3^B$	$+\phi_1^A\psi_1^{AB}Z_2(B,U)Q_{2,3}^{B,A,0}p_4^A\alpha_4^A$	$+\phi_1^A\psi_1^{AB}Z_2(B,U)Q_{2,3}^{B,B,0}p_4^B\alpha_4^B$	$+\phi_1^A\psi_1^{AB}Z_2(B,U)Q_{2,3}^{B,A,U}p_4^A\alpha_4^A$	$+\phi_1^A\psi_1^{AB}Z_2(B,U)Q_{2,3}^{B,B,U}p_4^B\alpha_4^B$	
$\mathbf{n}_{1,2}^{B,-,A}$	$\mathbf{n}_{1,2}^{B,-,B}$	$\mathbf{n}_{1,3}^{B,0,A}$ … $\mathbf{n}_{1,4}^{B,0U,B}$	$\mathbf{n}_{1,3}^{B,U,A}$	$\mathbf{n}_{13}^{B,U,B}$	$\mathbf{n}_{14}^{B,U0,A}$	$\mathbf{n}_{14}^{B,U0,B}$	$\mathbf{n}_{14}^{B,UU,A}$	$\mathbf{n}_{14}^{B,UU,B}$	v_1^B
$O_{1,1}^{B,A,-}$	$O_{1,1}^{B,B,-}$	$O_{1,2}^{A,A,0}$ … $O_{1,3}^{A,B,0U}$	$O_{1,2}^{B,A,U}$	$O_{1,2}^{B,B,U}$	$O_{1,3}^{B,A,U0}$	$O_{1,3}^{B,B,U0}$	$O_{1,3}^{B,A,UU}$	$O_{1,3}^{B,B,UU}$	$1-\lambda_1^B$
			$=Q_{1,2}^{B,A,U}p_3^A\alpha_3^A$	$=Q_{1,2}^{B,B,U}p_3^B\alpha_3^B$	$=Q_{1,3}^{B,A,U0}p_4^A\alpha_4^A$	$=Q_{1,3}^{B,B,U0}p_4^B\alpha_4^B$	$=Q_{1,3}^{B,A,UU}p_4^A\alpha_4^A$	$=Q_{1,3}^{B,B,UU}p_4^B\alpha_4^B$	
			$=\phi_1^B\psi_1^{BA}Z_2(A,U)Q_{2,2}^{A,A,-}p_3^A\alpha_3^A$	$=\phi_1^B\psi_1^{BA}Z_2(A,U)Q_{2,2}^{A,B,-}p_3^B\alpha_3^B$	$=\phi_1^B(\psi_1^{BA}Z_2(A,U)Q_{2,3}^{A,B,0}p_4^A\alpha_4^A$	$=\phi_1^B\psi_1^{BA}Z_2(A,U)Q_{2,3}^{A,B,0}p_4^B\alpha_4^B$	$=\phi_1^B(\psi_1^{BA}Z_2(A,U)Q_{2,3}^{A,A,U}p_4^A\alpha_4^A$	$=\phi_1^B(\psi_1^{BA}Z_2(A,U)\,Q_{2,3}^{A,B,U}p_4^B\alpha_4^B$	
			$+\phi_1^B\psi_1^{BB}Z_2(B,U)Q_{2,2}^{B,A,-}p_3^A\alpha_3^A$	$+\phi_1^B\psi_1^{BB}Z_2(B,U)Q_{2,2}^{B,B,-}p_3B\alpha_3^B$	$+\phi_1^B\psi_1^{BB}Z_2(B,U)Q_{2,3}^{B,A,0}p_4^A\alpha_4^A$	$+\phi_1^B\psi_1^{BB}Z_2(B,U)Q_{2,3}^{B,B,0}p_4^B\alpha_4^B$	$+\phi_1^B\psi_1^{BB}Z_2(B,U)Q_{2,3}^{B,A,U}p_4^A\alpha_4^A$	$+\phi_1^B\psi_1^{BB}Z_2(B,U)Q_{2,3}^{B,B,U}p_4^B\alpha_4^B$	
$\mathbf{w}_{12}^{U,-,A}$	$\mathbf{w}_{12}^{U,-,B}$	$\mathbf{w}_{13}^{U,0,A}$ … $\mathbf{w}_{14}^{U,0U,B}$	$\mathbf{w}_{13}^{U,U,A}$	$\mathbf{w}_{13}^{U,U,B}$	$\mathbf{w}_{14}^{U,U0,A}$	$\mathbf{w}_{14}^{U,U0,B}$	$\mathbf{w}_{14}^{U,UU,A}$	$\mathbf{w}_{14}^{U,UU,B}$	b_1^U
$\pi_1^A O_{1,1}^{A,A,-}$	$\pi_1^A O_{1,1}^{A,B,-}$	$\gamma_{1,3}^{U,A,0}$ … $\gamma_{1,4}^{U,B,0U}$	$\pi_1^A O_{1,2}^{A,A,U}$	$\pi_1^A O_{1,2}^{A,B,U}$	$\pi_1^A O_{1,3}^{A,A,U0}$	$\pi_1^A O_{1,3}^{A,B,U0}$	$\pi_1^A O_{1,3}^{A,A,UU}$	$\pi_1^A O_{1,3}^{A,B,UU}$	$1-\lambda_1^U$
$+\pi_1^B O_{1,1}^{B,A,-}$	$+\pi_1^B O_{1,1}^{B,B,-}$		$+\pi_1^B O_{1,2}^{B,A,U}$	$+\pi_1^B O_{1,2}^{B,B,U}$	$+\pi_1^B O_{1,3}^{B,A,U0}$	$+\pi_1^B O_{1,3}^{B,B,U0}$	$+\pi_1^B O_{1,3}^{B,A,UU}$	$+\pi_1^B O_{1,3}^{B,B,UU}$	
			$\mathbf{w}_{23}^{U,-,A}$	$\mathbf{w}_{23}^{U,-,B}$	$\mathbf{w}_{24}^{U,0,A}$	$\mathbf{w}_{24}^{U,0,B}$	$\mathbf{w}_{24}^{U,U,A}$	$\mathbf{w}_{24}^{U,U,B}$	b_2^U
-	-	- … -	$\pi_2^A O_{2,2}^{A,A,-}$	$\pi_2^A O_{2,2}^{A,B,-}$	$\pi_2^A O_{2,3}^{A,A,0}$	$\pi_2^A O_{2,3}^{A,B,0}$	$\pi_2^A O_{2,3}^{A,A,U}$	$\pi_2^A O_{2,3}^{A,B,U}$	$1-\lambda_2^U$
-	-	- … -	$+\pi_B O_{2,2}^{B,A,-}$	$+\pi_2^B O_{2,2}^{B,B,-}$	$+\pi_2^B O_{2,3}^{B,A,0}$	$+\pi_2^B O_{2,3}^{B,B,0}$	$+\pi_2^B O_{2,3}^{B,A,U}$	$+\pi_2^B O_{2,3}^{B,B,U}$	
-	-	- … -	$\mathbf{n}_{23}^{A,-,A}$	$\mathbf{n}_{23}^{A,-,B}$	$\mathbf{n}_{24}^{A0A}$	$\mathbf{n}_{24}^{A0B}$	$\mathbf{n}_{24}^{AUA}$	$\mathbf{n}_{24}^{AUB}$	v_2^A
-	-	- … -	$O_{2,2}^{A,A,-}$	$O_{2,2}^{A,B,-}$	$O_{2,3}^{A,A,0}$	$O_{2,3}^{A,B,0}$	$O_{2,3}^{A,A,U}$	$O_{2,3}^{A,B,U}$	$1-\lambda_2^A$
-	-	- … -	$=Q_{2,2}^{A,A,-}p_3^A\alpha_3^A$	$=Q_{2,2}^{A,B,-}p_3^B\alpha_3^B$	$=Q_{2,3}^{A,A,0}p_4^A\alpha_4^A$	$=Q_{2,3}^{A,B,0}p_4^B\alpha_4^B$	$=Q_{2,3}^{A,A,U}p_4^A\alpha_4^A$	$=Q_{2,3}^{A,B,U}p_4^B\alpha_4^B$	
-	-	- … -	$\mathbf{n}_{23}^{B,-,A}$	$\mathbf{n}_{23}^{B,-,B}$	$\mathbf{n}_{24}^{B,0,A}$	$\mathbf{n}_{24}^{B,0,B}$	$\mathbf{n}_{24}^{B,U,A}$	$\mathbf{n}_{24}^{B,U,B}$	v_2^B
-	-	- … -	$O_{2,2}^{B,A,-}$	$O_{2,2}^{B,B,-}$	$O_{2,3}^{B,A,0}$	$O_{2,3}^{B,B,0}$	$O_{2,3}^{B,A,U}$	$O_{2,3}^{B,B,U}$	$1-\lambda_2^B$
-	-	- … -	$=Q_{2,2}^{B,A,-}p_3^A\alpha_3^A$	$=Q_{2,2}^{B,B,-}p_3^B\alpha_3^B$	$=Q_{2,3}^{B,A,0}p_4^A\alpha_4^A$	$=Q_{2,3}^{B,B,0}p_4^B\alpha_4^B$	$=Q_{2,3}^{B,A,U}p_4^A\alpha_4^A$	$=Q_{2,3}^{B,B,U}p_4^B\alpha_4^B$	

observations made at occasion i, which is displayed in Table 4.6: the bases probabilities are denoted in red whilst the terms forming the mixing probabilities are denoted in blue. To further simplify the probability expressions, let $\zeta_{i,j}^{rs}$ denote the cell-probabilities of the conditional multinomials formed by the n-terms pooled over the possible partial histories $\sum_{z} n_{i,j}^{r,z_{i+1:j-1},s}$ (see Table 4.5). When needed, we add the element $h(t)$ to the superscript so as to specify the capture history element at time t, with $(i+1) \leq t \leq j$.

From Table 4.6, which presents the detailed expression of the cell-probabilities, we observe the following relationship between the $\zeta_{i-1,j}^{r,s,h(i)=U}$, cell-probabilities of the conditional multinomials corresponding to the animals released in a known state r at $i-1$, partially observed (code U) at i, and next seen in a known state s at j (rows M_1 to M_R in Table 4.6), and the cell-probabilities pertaining to the animals released at i in a known state l and next seen in a known state s at j (rows B_1 to B_R in Table 4.6), $\zeta_{i,j}^{l,s}/\lambda_i^l$:

$$\zeta_{i-1,t}^{rs,h(i)=U} = \sum_{l=1}^{R} \left(\frac{\left[\phi_{i-1}^{r} \psi_{i-1}^{r,l} Z_i(l,U)\right] \lambda_i^l}{\lambda_{i-1}^{r} - \sum_{g=1}^{R} O_{i-1,i-1}^{r,g,-} - \sum_{j=i+1}^{T} \sum_{z} \sum_{g=1}^{R} O_{i-1,j-1}^{r,g,z_{i:j-1},h(i)=0}} \right) \frac{\zeta_{i,t}^{ls}}{\lambda_i^l} \,. \tag{4.2}$$

Note that the quantity in blue does not depend on the time t or state s of first re-observation in a known state. Since

$$\lambda_{i-1}^{r} - \sum_{g=1}^{R} O_{i-1,i-1}^{r,g,-} - \sum_{j=i+1}^{T} \sum_{z} \sum_{g=1}^{R} O_{i-1,j-1}^{r,g,z_{i:j-1},h(i)=0} = \sum_{t=i+1}^{T} \sum_{g=1}^{R} \sum_{l=1}^{R} \left[\phi_{i-1}^{r} \psi_{i-1}^{r,l} Z_i(l,U)\right] \zeta_{i,t}^{lg} \,,$$

and that $\lambda_i^l = \sum_{t=i+1}^{T} \sum_{g=1}^{R} \zeta_{i,t}^{lg}$, it follows that

$$\sum_{l=1}^{R} \left(\frac{\left[\phi_{i-1}^{r} \psi_{i-1}^{r,l} Z_i(l,U)\right] \lambda_i^l}{\lambda_{i-1}^{r} - \sum_{g=1}^{R} O_{i-1,i-1}^{r,g,-} - \sum_{j=i+1}^{T} \sum_{z} \sum_{g=1}^{R} O_{i-1,j-1}^{r,g,z_{i:j-1},h(i)=0}} \right) = 1 \,. \tag{4.3}$$

Thus, we have shown that the number of animals previously released in a

known state r at $i-1$, partially observed at occasion i and re-observed later in a known state, follows a conditional multinomial distribution, which is a mixture of the conditional multinomial distributions followed by the animals released at i in the observable states, with the mixing probabilities equal to:

$$\frac{\left[\phi_{i-1}^{r}\psi_{i-1}^{r,l}Z_i(l,U)\right]\lambda_i^l}{\lambda_{i-1}^{r}-\sum_{g=1}^{R}O_{i-1,i-1}^{r,g}-\sum_{j=i+1}^{T}\sum_{z}\sum_{g=1}^{R}O_{i-1,j-1}^{r,g,z_{i:j-1},h(i)=0}} \text{ for } l=1,\ldots,R\,.$$

Going back to Table 4.6, this means that each row M_1 to M_R is a mixture of the bases formed by the rows B_1 to B_R.

Let $\zeta_{i,j}^{U,s}$ denote the conditional cell-probabilities associated with the animals first released at i in an unknown state and seen for the first time in a known state s at occasion j (i.e. the w-terms from Table 4.5 with associated cell-probabilities M_{R+2} from Table 4.6). The cell-probabilities corresponding to the w-terms associated with M_{R+1} are denoted by $\zeta_{i-1,j}^{U,s,h(i)=U}$. Then, for the animals first released at $i-1$ in an unknown state and seen in U at i:

$$\zeta_{i-1,t}^{U,s,h(i)=U}=\sum_{l=1}^{R}\left(\frac{\left[\sum_{v=1}^{R}\pi_{i-1}^{v}\phi_{i-1}^{v}\psi_{i-1}^{v,l}Z_i(l,U)\right]\lambda_i^l}{\lambda_{i-1}^{U}-\sum_{v=1}^{R}\pi_{i-1}^{v}\sum_{g=1}^{R}O_{i-1,i-1}^{v,g,-}-\sum_{j=i+1}^{T}\sum_{z}\sum_{g=1}^{R}\gamma_{i-1,j-1}^{U,g,z_{i:j-1},h(i)=0}}\right)\frac{\zeta_{i,t}^{l,s}}{\lambda_i^l}\,. \tag{4.4}$$

Once again, the quantity in blue does not depend on the time t or state s of first re-observation in a known state. As for the animals first observed in U at i,

$$\zeta_{i,t}^{U,s}=\sum_{l=1}^{R}\left(\frac{\pi_i^l\lambda_i^l}{\lambda_i^U}\right)\frac{\zeta_{i,t}^{l,s}}{\lambda_i^l}\,. \tag{4.5}$$

Using the same reasoning as above, it can be shown that $\sum_{l=1}^{R}\left(\frac{\pi_i^l\lambda_i^l}{\lambda_i^U}\right)=1$ and $\sum_{l=1}^{R}\left(\frac{\left[\sum_{v=1}^{R}\pi_{i-1}^{v}\phi_{i-1}^{v}\psi_{i-1}^{v,l}Z_i(l,U)\right]\lambda_i^l}{\lambda_{i-1}^{U}-\sum_{v=1}^{R}\pi_{i-1}^{v}\sum_{g=1}^{R}O_{i-1,i-1}^{v,g}-\sum_{j=i+1}^{T}\sum_{z}\sum_{g=1}^{R}\gamma_{i-1,j-1}^{U,g,z_{i:j-1},h(i)=0}}\right)=1$.

Therefore the numbers of animals first released at $i-1$ or i in an unknown state and first re-observed in a known state at a later occasion, both follow

a conditional multinomial distribution which is a mixture of the conditional multinomial distributions followed by the animals released at i in each of the R observable states, with the mixing probabilities respectively equal to:

$$\frac{\left[\sum_{v=1}^{R} \pi_{i-1}^{v} \phi_{i-1}^{v} \psi_{i-1}^{v,l} Z_i(l,U)\right] \lambda_i^l}{\lambda_{i-1}^U - \sum_{v=1}^{R} \pi_{i-1}^{v} \sum_{g=1}^{R} O_{i-1,i-1}^{v,g} - \sum_{j=i+1}^{T} \sum_z \sum_{g=1}^{R} \gamma_{i-1,j-1}^{U,g,z_{i:j-1},h(i)=0}} \text{ for } l = 1, \ldots, R$$

and

$$\frac{\pi_i^l \lambda_i^l}{\lambda_i^U} \text{ for } l = 1, \ldots, R$$

We have shown that the number of animals observed at i in an unknown state, released at $i-1$ in a known state or first released in U at i or $i-1$, and next re-observed in a known state all follow conditional multinomial distributions which are mixtures (with different mixing probabilities) of the same bases. These bases are formed by the animals seen at i in each of the observable states. The mixtures correspond to rows M_1 to M_{R+2} in Table 4.6, formed by the n-terms of animals released at $i-1$, and the w-terms of animals first released before or at i in Table 4.5; whilst the bases correspond to rows B_1 to B_R in Table 4.6 formed by the n-terms of animals released at i from Table 4.5.

More generally, animals that are partially observed at occasion i and re-observed in a known state afterwards can be grouped by their last observation in a known state before i, or, if they were never seen in a known state before i, by their time of first release in an unknown state. By conditioning on the state and time of their last observation in a known state or their time of first release in an unknown state, we obtain multinomial distributions corresponding to the time and state of the first re-observation in a known state after i. Hence, it can be shown that animals observed at i in an unknown state, released at any time t_1 before i in a known state or observed only in an unknown state before i, and re-observed in a known state afterwards, are also characterised

by the mixture property, with the bases being B_1 to B_R. Indeed, for animals released in a known state before i, the cell-probabilities of the corresponding multinomial can be expressed as follows.

Due to space constraints, let D^r be defined as $D^r = \sum_{g=1}^{R} \sum_{j=t_1+1}^{i} \sum_z O_{t_1,j-1}^{r,g,z_{t_1+1:j-1}} + \sum_{g=1}^{R} \sum_{j=i+1}^{T} \sum_z O_{t_1,j-1}^{r,g,z_{t_1+1:j-1},h(i)=0}$. The quantities forming D^r correspond respectively to the probability that an animal released in a known state r at time t_1 is re-observed in a known state before or at i and the probability that an animal is not observed at i but re-observed in a known state afterwards. (Recall that our test pertains only to the animals observed in U at i, all the terms corresponding to the animals not observed at this occasion are peeled off).

$$\zeta_{t_1,t>i}^{rs,h(i)=U} = \sum_{a=1}^{R} \left(\frac{\left[\sum_z Q_{t_1,i-1}^{r,a,z_{(t_1+1):(i-1)}} Z_i(a,U) \right] \lambda_i^a}{\lambda_{t_1}^r - D^r} \right) \frac{\zeta_{i,t}^{a,s}}{\lambda_i^a} \,. \tag{4.6}$$

Since

$$\lambda_{t_1}^r - D^r = \sum_{t=i+1}^{T} \sum_{g=1}^{R} \sum_{a=1}^{R} \sum_z \left[Q_{t_1,i-1}^{r,a,z_{(t_1+1):(i-1)}} Z_i(a,U) \right] \zeta_{i,t}^{ag} \,,$$

and that $\lambda_i^a = \sum_{t=i+1}^{T} \sum_{g=1}^{R} \zeta_{i,t}^{ag}$, it follows that

$$\sum_{a=1}^{R} \left(\frac{\left[\sum_z Q_{t_1,i-1}^{r,a,z_{(t_1+1):(i-1)}} Z_i(a,U) \right] \lambda_i^a}{\lambda_{t_1}^r - D^r} \right) = 1 \,. \tag{4.7}$$

Thus, we have shown that the number of animals previously released in a known state r, partially observed at occasion i and re-observed later in a known state, follows a conditional multinomial distribution, which is a mixture of the conditional multinomial distributions followed by the animals released

at i in the observable states, with the mixing probabilities equal to:

$$\left(\frac{\left[\sum_z Q_{t_1,i-1}^{r,a,z_{(t_1+1):(i-1)}} Z_i(a,U) \right] \lambda_i^a}{\lambda_{t_1}^r - D^r} \right) \text{ for } a = 1, \ldots, R\,.$$

In the same way, for the animals first released at a time $t_1 < i$ in an unknown state, we define D^U as

$$D^U = \sum_{g=1}^{R} \sum_{j=t_1+1}^{i} \sum_z \gamma_{t_1,j-1}^{r,g,z_{t_1+1:j-1}} + \sum_{g=1}^{R} \sum_{j=i+1}^{T} \sum_z \gamma_{t_1,j-1}^{r,g,z_{t_1+1:j-1},h(i)=0}$$

$$\zeta_{t_1,t>i}^{U,s,h(i)=U} = \sum_{a=1}^{R} \left(\frac{\left[\sum_{v=1}^{R} \pi_{t_1}^v \sum_z Q_{t_1,i-1}^{v,a,z_{(t_1+1):i-1}} Z_i(a,U) \right] \lambda_i^a}{\lambda_{t_1}^U - D^U} \right) \frac{\zeta_{i,t}^{a,s}}{\lambda_i^a}\,. \tag{4.8}$$

In the same manner as before,

$$\lambda_{t_1}^U - D^U = \sum_{a=1}^{R} \sum_{t=i+1}^{T} \sum_{g=1}^{R} \sum_{v=1}^{R} \pi_{t_1}^v \sum_z Q_{t_1,i-1}^{v,a,z_{(t_1+1):i-1}} Z_i(a,U) \zeta_{i,t}^{a,g}\,,$$

Hence,

$$\left(\frac{\left[\sum_{v=1}^{R} \pi_{t_1}^v \sum_z Q_{t_1,i-1}^{v,a,z_{(t_1+1):i-1}} Z_i(a,U) \right] \lambda_i^a}{\lambda_{t_1}^U - D^U} \right) = 1\,. \tag{4.9}$$

We have shown that the number of animals first released before i in an unknown state, partially observed at occasion i and re-observed later in a known state, follows a conditional multinomial distribution, which is a mixture of the conditional multinomial distributions followed by the animals released at i in the observable states, with the mixing probabilities equal to:

$$\left(\frac{\left[\sum_{v=1}^{R} \pi_{t_1}^v \sum_z Q_{t_1,i-1}^{v,a,z_{(t_1+1):i-1}} Z_i(a,U) \right] \lambda_i^a}{\lambda_{t_1}^U - D^U} \right) \text{ for } a = 1, \ldots, R\,.$$

Finally, using the following property cited from Pradel et al. (2003): "*if* $B1$

Table 4.5: Animals observed at i and later re-observed in a known state, previously released at i or $i-1$ in a known state, or first released at i or $i-1$ in an unknown state: sufficient statistic terms, for a capture-recapture experiment with T sampling occasions and R observable states ranging from A to R. The mixtures are denoted in blue and the bases in red.

$j=$	$i+1$				T			
$s=$	A	$\dots$	R	$\dots$	A	$\dots$	R	
	$\mathrm{n}_{i-1,i+1}^{A,U,A}$	$\dots$	$\mathrm{n}_{i-1,i+1}^{A,U,R}$	$\dots$	$\sum_z n_{i-1,T}^{A,z_{i:T-1},h(i)=U,A}$	$\dots$	$\sum_z n_{i-1,T}^{A,z_{i:T-1},h(i)=U,R}$	
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		
	$\mathrm{n}_{i-1,i+1}^{R,U,A}$	$\dots$	$\mathrm{n}_{i-1,i+1}^{R,U,R}$	$\dots$	$\sum_z n_{i-1,T}^{R,z_{i:T-1},h(i)=U,A}$	$\dots$	$\sum_z n_{i-1,T}^{R,z_{i:T-1},h(i)=U,R}$	
	$\mathrm{w}_{i-1,i+1}^{U,U,A}$	$\dots$	$\mathrm{w}_{i-1,i+1}^{U,U,R}$	$\dots$	$\sum_z w_{i-1,T}^{U,z_{i:T-1},h(i)=U,A}$	$\dots$	$\sum_z w_{i-1,T}^{U,z_{i:T-1},h(i)=U,R}$	
	$\mathrm{w}_{i,i+1}^{U,\cdot,A}$	$\dots$	$\mathrm{w}_{i,i+1}^{U,\cdot,R}$	$\dots$	$\sum_z w_{i,T}^{U,z_{i+1:T-1},A}$	$\dots$	$\sum_z w_{i,T}^{U,z_{i+1:T-1},R}$	
	$\mathrm{n}_{i,i+1}^{A,\cdot,A}$	$\dots$	$\mathrm{n}_{i,i+1}^{A,\cdot,R}$	$\dots$	$\sum_z n_{i,T}^{A,z_{i+1:T-1},A}$	$\dots$	$\sum_z n_{i,T}^{A,z_{i+1:T-1},R}$	
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$			
	$\mathrm{n}_{i,i+1}^{R,\cdot,A}$	$\dots$	$\mathrm{n}_{i,i+1}^{R,\cdot,R}$	$\dots$	$\sum_z n_{i,T}^{R,z_{i+1:T-1},A}$	$\dots$	$\sum_z n_{i,T}^{R,z_{i+1:T-1},R}$	

and $B2$ are mutually independent stochastic vectors, which are multinomially distributed, and if $M1$ and $M2$ are mutually independent stochastic vectors whose distributions are separately mixtures of the distributions of $B1$ and $B2$, then the distribution of $M1+M2$ is itself a mixture of the distributions of $B1$ and $B2$", the conditional multinomials of the animals released in a known state or first released in an unknown state before or at i, and partially observed at i can be pooled in the same manner as for Test M as shown in Table 4.7. Thus, the table used to test the mixture property of partial observations at occasion i is given in Table 4.7: the bases are denoted in red and the mixtures in blue. The MMLM approach is then used:

- Estimation of the cell-probabilities of the bases and of the mixing probabilities via maximum-likelihood,
- Goodness-of-fit assessment based on the distance between expected and observed values.

We implement this procedure by building on the code developed in Matlab

Table 4.6: Cell-probabilities of the conditional multinomials of Table 4.5. The terms constitutive of the mixtures are denoted in blue and those constitutive of the bases in red.

$j=$	$i+1$	...	R	...	T	...	R	
$s=$	A	...	R	...	A	...	R	
M_1	$\phi_{i-1}^A \sum_{l=1}^R \psi_{i-1}^{A,l} Z_i(l,U) Q_{i,i}^{l,A,-} p_{i+1}^A \alpha_{i+1}^A$	...	$\phi_{i-1}^A \sum_{l=1}^R \psi_{i-1}^{A,l} Z_i(l,U) Q_{i,i}^{l,R,-} p_{i+1}^R \alpha_{i+1}^R$	...	$\phi_{i-1}^A \sum_{l=1}^R \psi_{i-1}^{A,l} Z_i(l,U) \sum_z Q_{i,T-1}^{l,A,z_{i+1:T-1}} p_T^A \alpha_T^A$	...	$\phi_{i-1}^A \sum_{l=1}^R \psi_{i-1}^{A,l} Z_i(l,U) \sum_z Q_{i,T-1}^{l,R,z_{i+1:T-1}} p_T^R \alpha_T^R$	$\times \frac{1}{\lambda_{i-1}^A - \sum_{g=1}^R O_{i-1,i-1}^{A,g,-} - \sum_{j=i+1}^T \sum_z \sum_{g=1}^R O_{i-1,j-1}^{A,g,z_{i:j-1},h(i)=0}}$
M_R	$\phi_{i-1}^R \sum_{l=1}^R \psi_{i-1}^{R,l} Z_i(l,U) Q_{i,i}^{l,A,-} p_{i+1}^A \alpha_{i+1}^A$	...	$\phi_{i-1}^R \sum_{l=1}^R \psi_{i-1}^{R,l} Z_i(l,U) Q_{i,i}^{l,R,-} p_{i+1}^R \alpha_{i+1}^R$	...	$\phi_{i-1}^R \sum_{l=1}^R \psi_{i-1}^{R,l} Z_i(l,U) \sum_z Q_{i,T-1}^{l,A,z_{i+1:T-1}} p_T^A \alpha_T^A$	...	$\phi_{i-1}^R \sum_{l=1}^R \psi_{i-1}^{R,l} Z_i(l,hU(i)) \sum_z Q_{i,T-1}^{l,R,z_{i+1:T-1}} p_T^R \alpha_T^R$	$\times \frac{1}{\lambda_{i-1}^R - \sum_{g=1}^R O_{i-1,i-1}^{R,g,-} - \sum_{j=i+1}^T \sum_z \sum_{g=1}^R O_{i-1,j-1}^{R,g,z_{i:j-1},h(i)=0}}$
M_{R+1}	$\sum_{v=1}^R \pi_{i-1}^v \phi_{i-1}^v \sum_{l=1}^R \psi_{i-1}^{v,l} Z_i(l,U) Q_{i,i}^{l,A,-} p_{i+1}^A \alpha_{i+1}^A$	...	$\sum_{v=1}^R \pi_{i-1}^v \phi_{i-1}^v \sum_{l=1}^R \psi_{i-1}^{v,l} Z_i(l,U) Q_{i,i}^{l,R,-} p_{i+1}^R \alpha_{i+1}^R$	...	$\sum_{v=1}^R \pi_{i-1}^v \phi_{i-1}^v \sum_{l=1}^R \psi_{i-1}^{v,l} Z_i(l,U) \sum_z Q_{i,T-1}^{l,A,z_{i+1:T-1}} p_T^A \alpha_T^A$	...	$\sum_{v=1}^R \pi_{i-1}^v \phi_{i-1}^v \sum_{l=1}^R \psi_{i-1}^{v,l} Z_i(l,U) \sum_z Q_{i,T-1}^{l,R,z_{i+1:T-1}} p_T^R \alpha_T^R$	$\times \frac{1}{\lambda_{i-1}^U - \sum_{v=1}^R \pi_{i-1}^v \sum_{g=1}^R O_{i-1,i-1}^{v,g,-} - \sum_{j=i+1}^T \sum_z \sum_{g=1}^R \gamma_{i-1,j-1}^{U,g,z_{i:j-1},h(i)=0}}$
M_{R+2}	$\sum_{v=1}^R \pi_i^v \, Q_{i,i}^{v,A,-} p_{i+1}^A \alpha_{i+1}^A$	...	$\sum_{v=1}^R \pi_i^v \, (Q_{i,i}^{v,R,-} p_{i+1}^R \alpha_{i+1}^R)$	...	$\sum_{v=1}^R \pi_i^v \sum_z Q_{i,T-1}^{v,A,z_{i+1:T-1}} p_T^A \alpha_T^A$	...	$\sum_{v=1}^R \pi_i^v \sum_z Q_{i,T-1}^{v,R,z_{i+1:T-1}} p_T^R \alpha_T^R$	$\times \frac{1}{\lambda_i^U}$
B_1	$Q_{i,i}^{A,A,-} p_{i+1}^A \alpha_{i+1}^A$	...	$Q_{i,i}^{A,R,-} p_{i+1}^R \alpha_{i+1}^R$	...	$\sum_z Q_{i,T-1}^{A,A,z_{i+1:T-1}} p_T^A \alpha_T^A$	...	$\sum_z Q_{i,T-1}^{A,R,z_{i+1:T-1}} p_T^R \alpha_T^R$	$\times \frac{1}{\lambda_i^A}$
B_R	$Q_{i,i}^{R,A,-} p_{i+1}^A \alpha_{i+1}^A$	...	$Q_{i,i}^{R,R,-} p_{i+1}^R \alpha_{i+1}^R$	...	$\sum_z Q_{i,T-1}^{R,A,z_{i+1:T-1}} p_T^A \alpha_T^A$	...	$\sum_z Q_{i,T-1}^{R,R,z_{i+1:T-1}} p_T^R \alpha_T^R$	$\times \frac{1}{\lambda_i^R}$

by R. Choquet, for the existing Test M as implemented in program U-CARE. Note that the parameters estimated here are not the capture-recapture parameters presented in Table 4.6, but the simpler terms presented in Table 4.2, since we only need to assess the mixture property demonstrated in this section.

4.3.2 Goodness-of-fit

For the goodness-of-fit assessment, we consider different statistics based on the distance between expected values under the model and observed values: Pearson's χ^2, the log-likelihood ratio statistic G^2 (Cressie and Read, 1988, p. 10); and more generally, due to the different properties of these statistics depending on the alternatives or sparseness of the table, the power-divergence family of statistics, which we denote $CR(\lambda)$ (Cressie and Read, 1988).
Let O_k and E_k respectively denote the observed and expected frequencies. The test statistics are defined by:

$$\chi^2 = \sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k} \tag{4.10}$$

$$G^2 = 2 \sum_{k=1}^{K} O_k \ \ln\left(\frac{O_k}{E_k}\right) \tag{4.11}$$

$$CR(\lambda) = \frac{2}{\lambda(\lambda+1)} \sum_{k=1}^{K} O_k \left(\left[\frac{O_k}{E_k}\right]^{\lambda} - 1\right) \tag{4.12}$$

More specifically, we use the power-divergence statistic with $\lambda = 2/3$, which is recommended by Cressie and Read (1988) for its properties as a compromise between X^2 and G^2. We note that the G^2 and χ^2 both belong to the power-divergence family. Indeed,

$$CR(1) = \sum_{k=1}^{K} O_k \frac{O_k - E_k}{E_k} = \sum_{k=1}^{K} \frac{O_k^2}{E_k} - \sum_{k=1}^{K} O_k$$

Table 4.7: Table used for testing the mixture property of partial observations at occasion i. The columns are pooled over the different partial histories, $h(i) = U$ denotes that the animals are seen in U at i. The rows are pooled by state at last release (first R rows) and when there are no certain observations prior to i+1 (row R+1).

$j =$	$i+1$			$\dots$	T		
$s =$	A	$\dots$	R	$\dots$	A	$\dots$	R
	$\sum_{f=1}^{i-1} n_{f,i+1}^{A,.,A,h(i)=U}$	$\dots$	$\sum_{f=1}^{i-1} n_{f,i+1}^{A,.,R,h(i)=U}$	$\dots$	$\sum_{f=1}^{i-1} n_{f,T}^{A,.,A,h(i)=U}$	$\dots$	$\sum_{f=1}^{i-1} n_{f,T}^{A,.,R,h(i)=U}$
	$\vdots$		$\vdots$		$\vdots$		$\vdots$
	$\sum_{f=1}^{i-1} n_{f,i+1}^{R,.,A,h(i)=U}$	$\dots$	$\sum_{f=1}^{i-1} n_{f,i+1}^{R,.,R,h(i)=U}$	$\dots$	$\sum_{f=1}^{i-1} n_{f,T}^{R,.,A,h(i)=U}$	$\dots$	$\sum_{f=1}^{i-1} n_{f,T}^{R,.,R,h(i)=U}$
	$\sum_{f=1}^{i} w_{f,i+1}^{U,.,A,h(i)=U}$	$\dots$	$\sum_{f=1}^{i} w_{f,i+1}^{U,.,R,h(i)=U}$	$\dots$	$\sum_{f=1}^{i} w_{f,T}^{U,.,A,h(i)=U}$	$\dots$	$\sum_{f=1}^{i} w_{f,T}^{U,.,R,h(i)=U}$
	$n_{i,i+1}^{A,A}$	$\dots$	$n_{i,i+1}^{A,R}$	$\dots$	$n_{i,T}^{A,.,A}$	$\dots$	$n_{i,T}^{A,.,R}$
	$\vdots$		$\vdots$		$\vdots$		$\vdots$
	$n_{i,i+1}^{R,A}$	$\dots$	$n_{i,i+1}^{R,R}$	$\dots$	$n_{i,T}^{R,.,A}$	$\dots$	$n_{i,T}^{R,.,R}$

and

$$\chi^2 = \sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k} = \sum_{k=1}^{K} \frac{O_k^2}{E_k} - 2\sum_{k=1}^{K} O_k + \sum_{k=1}^{k} E_k \,.$$

Since the totals are fixed, $\sum_{i=1}^{k} E_k = \sum_{i=1}^{k} O_k$, and the two expressions are equivalent.

For G^2, let $y = O_k/E_k$, $y^\lambda = \exp[\lambda ln(y)]$. Using the usual Taylor series expansion at zero,

$$\exp[\lambda ln(y)] = \ln(y) + \frac{\lambda^2 \ln(y)^2}{2!} + \mathcal{O}(\{\lambda \ln(y)\}^{n+1}) \,.$$

Therefore, applying the Box-Cox transformation results in

$$\frac{y^\lambda - 1}{\lambda} = \ln(y) + \frac{\lambda \ln(y)^2}{2!} + \mathcal{O}(\{\lambda \ln(y)\}^{n+1}) \,.$$

When $\lambda \to 0$, all the terms in the expansion become negligible except for $\ln(y)$. Hence,

$$\lim_{\lambda \to 0} \frac{2}{\lambda(\lambda+1)} \sum_{k=1}^{K} O_k \left(\left[\frac{O_k}{E_k} \right]^{\lambda} - 1 \right) = \sum_{k=1}^{K} 2 O_k \ln \frac{O_k}{E_k} = G^2 \,.$$

If the null hypothesis is extrinsic, or in other words, if the parameter values are pre-specified under the null hypothesis, (Sokal and Rohlf, 2012, p. 711), these different power-divergence statistics are asymptotically equivalent and follow a χ^2 distribution with $K-1$ degrees of freedom (Cressie and Read, 1984). However, when the null hypothesis is intrinsic, i.e. the parameter values have to be estimated from the data, this approximation becomes conservative since the estimation artificially makes the expected values closer to the observed data, thus the null hypothesis is less likely to be rejected (Conover, 1980, p. 194). To remedy this, when the parameters are estimated via maximum likelihood of the multinomial cell frequencies, as is the case for the first step of our mixture test, one degree of freedom is removed per parameter estimated (Moore, 1986, p. 66). In order for the asymptotic distributions to hold, expected frequencies in each cell should be at least 2 for a level $\alpha = 0.05$ (Moore, 1986, p. 71), this is also the threshold used in U-CARE. As a practical way of checking whether the asymptotics hold, Reise and Revicki (2014, p. 112) advise comparing the results from different statistics that are supposed to be asymptotically equivalent. If the results are very different, it is likely that the asymptotics do not hold. In our simulations and applications, we check this based on the three goodness-of-fit statistics presented above. In addition to this, we note that the tables used at each occasion i condition on known states. Therefore, like for Test M, the test-statistics obtained at each occasion are independent and a global test-statistic can be computed by summing these up. This global test-statistic follows, under the null hypothesis, a chi-square distribution with the number of degrees of freedom being the sum of the degrees of freedom of the components per occasion.

4.4 Simulation results

Under the null hypothesis, animals partially observed at i and re-observed later are consistent with being a mixture of animals observed in the directly observable states at i and re-observed in the same conditions: the partial observations are generated solely by the observable states. Using the usual H_0 notation for the null hypothesis and H_1 for the alternative, $H_1 = \bar{H}_0$. A large array of situations come under the alternative hypothesis: from the partial observations being generated by the directly observable states *and* another state which is never directly observable to the more extreme case of partial observations all being generated only by one (or more) states which are never directly observable. We wish the test to be sensitive to all departures from the null hypothesis. In order to minimise the chances of sparse data and verify that the test works as expected in theory, we use simulation in very large sample size conditions and focus on an extreme case of the alternative hypothesis. First, we simulate capture-recapture data under the null hypothesis, arising from two directly observable states, with $K = 5$ sampling occasions and 25,000 animals released per occasion. The capture, survival and transition probabilities, are respectively set as $p^A = p^B = 0.6$, $\phi^A = 0.6$, $\phi^B = 0.9$, $\psi^{AB} = 0.8$, $\psi^{BA} = 0.7$. This scenario is denoted by $2S$. In order to introduce partial observations, we set a varying percentage of the observations as missing completely at random (MCAR). We also simulate data under the alternative hypothesis (also with $K = 5$ sampling occasions and 25,000 animals released per occasion), corresponding to the situation illustrated in Figure 4.2, where the partial observations are not generated by either of the two directly observable states, but by a third state which is never directly observable, this scenario is denoted by $3S$. Using the standard multievent notation introduced in Section 3.6 from

Chapter 3, the survival matrix is $\mathbf{\Phi_t} = \begin{bmatrix} 0.7 & 0 & 0 & 0.3 \\ 0 & 0.8 & 0 & 0.2 \\ 0 & 0 & 0.9 & 0.1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ for $t = 1, \ldots, 4$; the transition matrix is $\mathbf{\Psi_t} = \begin{bmatrix} 0.1 & 0.3 & 0.6 & 0 \\ 0.3 & 0.15 & 0.55 & 0 \\ 0.4 & 0.4 & 0.2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ for $t = 1, \ldots, 4$ and the event matrix $\mathbf{B_t} = \begin{bmatrix} 0.45 & 0.55 & 0 & 0 \\ 0.45 & 0 & 0.55 & 0 \\ 0.45 & 0 & 0 & 0.55 \\ 1 & 0 & 0 & 0 \end{bmatrix}$ for $t = 1, \ldots, 5$

We simulate 600 datasets for each scenario. If any of the expected values are lower than two, the corresponding test is deemed Non Applicable (NA). The results obtained are given in terms of percentage of significant test results out of the number of applicable tests, at a 5% level, in Table 4.8. The simulation results show that for the datasets simulated under the null hypothesis (scenario 2S), the Type I error rate is close to 5%, whatever the percentage of partial observations (apart from the extreme case of 95% of MCAR data). The results from the three statistics are very similar; this supports the fact that the theoretical asymptotics hold. Also, for the datasets simulated under an alternative hypothesis (scenario 3S), 100% of the test results were significant, whatever the test-statistic considered. Obviously, this very high power was to be expected due to the very large sample size considered. But, more importantly, it shows that the test reacts as expected from the theory of the previous section, when the partial observations are not generated by the directly observable states.

Table 4.8: Testing the mixture property of partial observations: simulation results, percentage of significant test results out of the number of applicable tests, G denotes the global test, i the sampling occasion and $\%MCAR$ the percentage of observations set to "Unknown" and N denotes the number of applicable tests.

Scenario	% MCAR	i	χ^2(%)	CR ($\lambda = \frac{2}{3}$)(%)	G^2(%)	*dof*	N (non NA)
2S	5	2	3.50	3.83	5.00	12	600
		3	5.83	5.67	6.50	6	600
		G	5.00	5.17	5.67	18	600
	15	2	4.50	4.17	4.67	12	600
		3	3.83	3.67	3.83	6	600
		G	4.17	4.00	4.00	18	600
	25	2	5.83	5.83	5.67	12	600
		3	3.83	4.00	4.00	6	600
		G	5.17	5.17	4.83	18	600
	35	2	6.00	5.83	5.83	12	600
		3	3.67	3.67	3.83	6	600
		G	6.33	6.17	6.67	18	600
	45	2	5.00	5.00	5.33	12	600
		3	6.17	6.17	6.17	6	600
		G	6.33	6.33	6.17	18	600
	55	2	5.67	6.00	5.83	12	600
		3	5.67	5.50	5.50	6	600
		G	6.00	5.67	5.33	18	600
	65	2	5.50	5.50	5.33	12	600
		3	5.83	5.83	5.50	6	600
		G	5.17	5.17	5.67	18	600
	75	2	4.83	5.17	5.83	12	600
		3	4.17	4.17	3.83	6	600
		G	5.17	5.17	6.00	18	600
	85	2	4.83	4.83	5.33	12	600
		3	5.33	5.17	5.50	6	600
		G	4.83	5.17	5.83	18	600
	95	2	NA	NA	NA	12	3[1]
		3	3.34	3.56	4.01	6	449[2]
		G	3.36	3.58	4.04	6	446[2]
3S	-	2	**100.00**	**100.00**	**100.00**	12	600
	-	3	**100.00**	**100.00**	**100.00**	6	600
	-	G	**100.00**	**100.00**	**100.00**	18	600

[1] results set to NA due to small number of applicable datasets
[2] number of datasets resulting in the same number of dof

4.5 Applications

We have shown that, in theory, our test has the ability to assess whether partial observations can be adequately modelled as stemming solely from the directly observable states in a capture-recapture experiment. In this section, we apply our new tool to real-life situations, for which the adequate underlying state structure is known, in order to determine whether our test can also have good performance for datasets with more realistic sample sizes. We use two datasets as applications: the Canada geese presented in Section 3.6 and a dataset of greater flamingoes. Note that the tables needed for the test were quite sparse, we therefore used the following pooling strategy: on the columns, pooled to the maximum until there was one degree of freedom left for the test (the column with the minimal sum is pooled with the column with the second minimal sum and so on) whilst on the rows, all the rows corresponding to mixtures are pooled so that there is just one mixture left to test for.

4.5.1 Canada geese

We examine the Canada geese dataset under both the null and alternative hypotheses by artificially creating these situations within the data. We set a varying percentage of the observed geese's states to unknown (MCAR): 15%, 25% and 45%, so that the partial observations are generated only by the observable states (H_0). These situations are respectively denoted by P15, P25 and P45 in Table 4.9. Then we examine situations that come under the alternative hypotheses by setting all of the observations from a particular state to "unknown". Hence, for situations 2PO and 3PO the partial observations stem, respectively, only from states 2 and 3. Note that the state set to "unknown" is never directly observed in each of these situations.

The results obtained from applying the mixture test to all these configurations of the geese dataset are given in Table 4.9, in terms of p-value for the

different test-statistics that were considered in this chapter. These results are very promising with the test reacting as it should under the different configurations examined. Under all the null hypothesis configurations, the directly observable states as sole underlying states for the partial observations provide an adequate fit to the data. The p-values obtained for the three test-statistics are extremely close, thus the asymptotics hold in these situations. For the configurations under the alternative, the test strongly rejects the null hypothesis, with $p<0.001$ for almost all the tests examined (by occasion and global). Hence, the results from configurations 2PO and 3 PO lead to the conclusion that the directly observable states do not provide an adequate underlying state-structure for the partial observations.

4.5.2 Greater flamingoes

The dataset of greater flamingoes is formed of 7061 individuals ringed and monitored from 1977 to 2014. It is a subset of the large flamingo study that has been conducted in the Camargue (France) since 1977 (Sanz-Aguilar et al., 2012, for example) and was provided by the Tour du Valat. Our dataset focusses on the breeding status of the flamingoes: at each occasion, they are observed as "breeder" or "non-breeder"; for some individuals, the state cannot be determined but is known to be either "non-breeder" or "failed breeder". This situation falls under the alternative hypothesis for our test, since the partial observations are not generated solely from the two directly observable states. Note however, that this setting is less extreme than the 3S scenario from our simulations in Section 4.4 or configurations 2PO and 3PO from the geese analyses. Indeed, the underlying state structure for the partial observations consist here in one of the directly observable states and another state which is not observable directly. The results obtained are shown in Table 4.10 in terms of p-values for the test-statistics considered, again, the tests per occasion are

Table 4.9: Using different configurations of the Canada geese dataset to assess the new mixture test for assessing the underlying state structure of partial observations, in some real-life conditions: the p-value obtained at each occasion is presented for the different statistics and the associated global tests are denoted by G.

Configuration	i	$\chi^2(\%)$	CR ($\lambda = \frac{2}{3}$)(%)	$G^2(\%)$	dof
P15	2	0.14	0.14	0.14	1
	3	0.14	0.14	0.14	1
	4	0.60	0.60	0.60	1
	G	0.21	0.20	0.20	3
P25	2	0.57	0.57	0.57	1
	3	0.09	0.09	0.09	1
	4	0.85	0.85	0.85	1
	G	0.35	0.35	0.34	3
P45	2	0.84	0.84	0.84	1
	3	0.82	0.82	0.82	1
	4	0.85	0.85	0.85	1
	G	0.99	0.99	0.99	3
2PO	2	**<0.001**	**<0.001**	**<0.001**	1
	3	**<0.001**	**<0.001**	**<0.001**	1
	4	**<0.001**	**<0.001**	**<0.001**	1
	G	**<0.001**	**<0.001**	**<0.001**	3
3PO	2	0.13	0.13	0.14	1
	3	**<0.001**	**<0.001**	**<0.001**	1
	4	**<0.001**	**<0.001**	**<0.001**	1
	G	**<0.001**	**<0.001**	**<0.001**	3

displayed as well as the pooled global test, denoted by G.

In this case, the test does not give the expected conclusion: its result is significant only at occasion 30; there is otherwise no evidence to reject H_0, with none of the other tests per occasion giving significant results and the global test resulting in a p-value around 0.27. Since we know that the dataset corresponds to a situation that belongs to the alternative hypothesis, we decided to use non-parametric bootstrap in order to empirically evaluate the power of the test for matched datasets. We generated 6,000 bootstrap samples of 7061 individuals and 38 occasions, by sampling 7061 individual capture histories from the original dataset with replacement.

The results obtained from the non-parametric bootstrap samples are presented in Table 4.11, in terms of percentage of significant test results. For the global test, around 99% of the results are significant: notice that more than 50% of the results for the tests per occasion are significant at occasions 11 and 30 (which was the only significant result for our original dataset). The bootstrapped results lead to the conclusion that the partial observations do not solely result from the observable states, for the flamingo data. As a counter-example, we also performed non-parametric bootstrap on configuration P45 from the geese dataset, which is under the null hypothesis, around 7.8% of the global test results were found to be significant at a 5% level, which is close to the Type I error rate, as would be expected.

As for the non-significant result obtained for the original flamingo dataset despite the high estimated power from the bootstrapped samples (based on the global test for instance), we performed simple descriptive statistics on the bootstrapped datasets, and these did not show any obvious discrepancies compared to the original flamingo dataset; further investigations were needed. We checked that there was no coding issue by implementing the bootstrap using two different computer packages (R and MATLAB); both implementations led to similar results. Upon examining bootstrapped datasets individually,

we noticed that there were very often a couple of significant results per occasion among the 34 occasions considered. These were enough to render the global test result significant, as it is formed by summing up the chi-square statistics obtained per occasion. Hence, the 99% of significant global test results obtained for the bootstrap. We also examined individual occasions for which there was a strong discrepancy between the test results from the original dataset and those from the bootstrap samples. For example, at occasion 11 ($p = 0.89$ versus 52% of significant results), the number of animals forming the mixture (i.e. partially observed animals) is quite small (N=44) and many of the bootstrapped datasets result in multinomial cells with very small frequencies (see Table 4.12). These are associated to extremely small cell probabilities which are much more extreme than those of the bases and cannot be reconstructed from a weighted average of the bases cell-probabilities; the test for a mixture structure is consequently rejected. More generally, the bootstrapped datasets are obtained by sampling with replacement and the over or under-representations within the cells of the mixture are not necessarily consistent with the over and under-representations within the cells of the bases. This is not surprising as the partial observations in the flamingo dataset are actually not generated by the mixture model assumed under the null hypothesis. When the cell-probabilities associated to the mixture are too extreme to be reconstructed from the bases, the test is rejected.

The result obtained from the original dataset could be a "false negative". The issue of reproducibility of the p-value might also be considered (see for example Boos and Stefanski, 2011; Halsey et al., 2015). Halsey et al. (2015) note that p-values are often used alone as an "index of truth" whilst, like other statistical quantities, they present sample-to-sample variability, which is rarely reported. To take this into consideration, Boos and Stefanski (2011) advise using variability measures such as bootstrap standard-errors or prediction intervals, or computing reproducibility probabilities. Note however, that

these criticisms are mainly targeted at non-reproducible significant p-values. Also, in the goodness-of-fit context we use the p-value as guidance rather than a truth indicator. However, computing and reporting p-value variability in these goodness-of-fit situations or routinely using non-parametric bootstrap, which do not require much computational power, could be of interest, and this is an area that could be explored more in future research.

4.6 Discussion

In this chapter, we have derived a mixture test that assesses whether partial observations are generated solely by the directly observable states. This test is based on distributional properties which we have demonstrated. It was shown to perform well in theory, based on simulations in very large sample size conditions. In order to assess the usefulness of the test in real-life situations, its power should be examined for smaller, more realistic, sample sizes. For example, simulations based on an incrementally decreasing sample size could be run so as to approximately evaluate the minimum sample size necessary for the test to work well; this is an area of future research.

As evidenced by our application on the greater flamingoes data, there is a range of alternatives that can be considered: partial observations may stem from one of the directly observable states and an additional state. Going further, they may also stem from all the observable states and another state which is never observable directly. In theory, the test should react to this situation too. However, in practice, we surmise that the other state would have to present different enough properties from the directly observable states for the test to be powerful enough to detect it. Evaluating the performance of the test for a range of alternatives could be explored in future research. Also, the issue of sample size is more complex than usual in this framework. Indeed it is not only the total sample size which matters but also the proportion of

Table 4.10: Testing the mixture property of partial observations on a dataset of greater flamingoes: p-value obtained at each occasion i and for the global test denoted by G.

i	$\chi^2(\%)$	CR ($\lambda = \frac{2}{3}$)(%)	$G^2(\%)$	dof
2	NA	NA	NA	NA
⋮	⋮	⋮	⋮	⋮
8	NA	NA	NA	NA
9	0.33	0.33	0.33	1
10	0.71	0.71	0.71	1
11	0.89	0.89	0.89	1
12	0.57	0.57	0.57	1
13	0.26	0.26	0.26	1
14	0.91	0.91	0.91	1
15	0.12	0.12	0.12	1
16	0.15	0.15	0.15	1
17	0.34	0.34	0.34	1
18	0.25	0.25	0.25	1
19	0.90	0.90	0.90	1
20	0.68	0.68	0.68	1
21	0.64	0.64	0.64	1
22	0.19	0.19	0.19	1
23	0.10	0.10	0.10	1
24	0.26	0.26	0.27	1
25	0.48	0.48	0.48	1
26	NA	NA	NA	NA
27	NA	NA	NA	NA
28	NA	NA	NA	NA
29	0.18	0.18	0.18	1
30	**0.03**	**0.03**	**0.03**	1
31	NA	NA	NA	NA
32	0.10	0.10	0.10	1
33	0.15	0.14	0.12	1
34	0.99	0.99	0.99	1
35	0.62	0.62	0.62	1
36	0.75	0.75	0.75	1
G	0.28	0.27	0.26	24

Table 4.11: Flamingo dataset, percentage of significant results for non-parametric bootstrap samples: tests per occasion and global test (denoted by G).

i	χ^2(%)	CR ($\lambda = \frac{2}{3}$)(%)	G^2(%)	N (non NA)
2	NA	NA	NA	0
⋮	⋮	⋮	⋮	⋮
6	NA	NA	NA	0
7	25.19	25.19	27.72	671
8	30.15	30.15	31.87	1048
9	14.05	14.05	14.14	5941
10	10.78	10.78	10.73	6000
11	**52.11**	**52.11**	**52.14**	5995
12	11.44	11.44	11.50	5998
13	27.83	27.83	27.85	6000
14	14.72	14.72	14.62	6000
15	44.88	44.88	44.82	6000
16	23.20	23.20	23.25	6000
17	14.54	14.54	14.55	5999
18	44.52	44.52	44.55	6000
19	33.58	33.58	33.58	6000
20	32.65	32.65	32.50	5966
21	10.13	10.13	10.12	6000
22	12.13	12.13	12.12	6000
23	25.05	25.05	24.98	6000
24	11.18	11.18	11.18	6000
25	15.48	15.48	15.45	6000
26	NA	NA	NA	0
27	NA	NA	NA	0
28	NA	NA	NA	0
29	20.80	20.80	20.80	5525
30	**67.32**	**67.32**	**67.28**	6000
31	NA	NA	NA	0
32	30.88	30.88	31.08	6000
33	37.57	37.57	39.74	5994
34	9.07	9.07	9.07	6000
35	11.37	11.37	11.32	6000
36	5.58	5.58	5.57	6000
G	**98.95**	**98.97**	**98.98**	6000

Table 4.12: Occasion 11, observed frequencies, observed probabilities and estimated probabilities under the null hypothesis (mixture structure) for the original flamingo dataset and an example bootstrapped dataset.

		Observed frequencies			Observed probabilities			Estimated probabilities		
Real dataset	Mixture	18	12	14	0.41	0.27	0.32	0.41	0.27	0.33
	Basis1	161	133	151	0.36	0.30	0.34	0.36	0.30	0.34
	Basis2	70	26	41	0.51	0.19	0.30	0.51	0.19	0.30
Bootstrapped	Mixture	32	3	10	0.71	0.07	0.22	0.50	0.26	0.23
	Basis1	168	150	124	0.38	0.34	0.28	0.38	0.34	0.28
	Basis2	48	39	27	0.42	0.34	0.24	0.50	0.26	0.23

partial observations.

Finally, regarding the interpretation of the test, if the null hypothesis is not rejected, the observable states provide an adequate underlying structure for the partial observations. However, the interpretation of a significant test result is not as straightforward. Indeed, if the observable states are inadequate, how many additional states should be considered for the underlying structure and how the partial observations should be modelled are questions that do not have obvious answers at this stage and constitute another area of future research.

Chapter 5

A procedure to test for unobservable states

5.1 Introduction

In this chapter, we focus on multistate models presenting an additional level of uncertainty for the non-captures, or in other words, some of the states are unobservable. Unobservable states form a general modelling tool. They can be used to model phenomena such as temporary emigration or to represent a biological state such as *non-breeder* for birds (Kendall, 2004), or *dormant* for plants (Kéry and Schaub, 2012). In this framework, when an individual is observed, there is no uncertainty as to its state. An example of a model structure containing an unobservable state, based on breeding status is illustrated in Figure 5.1. A bird may be observed as "Breeder" (B) or "Failed breeder" (FB); a bird may be a 'Non-breeder"(NB) but can never be observed in that state. Hence, when a bird is not captured (NC), it can actually be in any of the three live states: B, FB, or NB, or it can be dead (†).

Our initial goal was to develop a procedure to assess whether there is statistical support that one or more unobservable states should be defined in the

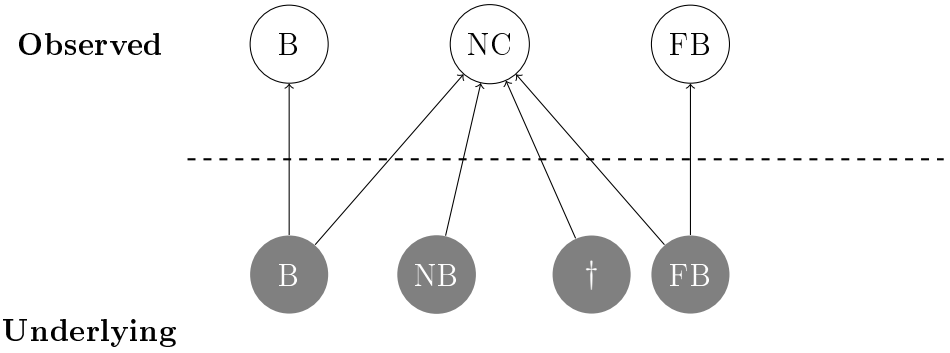


Figure 5.1: Illustration of a model structure with two observable states B and FB, one unobservable live state NB (and the dead state: †). The event NC denotes "Not captured "

model for it to fit the data adequately. Note that there is not necessarily a one-to-one correspondence between the unobservable states defined in the model and biological states. The general idea behind this procedure is a natural extension of the question posed by Test M (described in Section 4.2): at each occasion, are the animals that are not captured consistent with being a mixture of animals observed in any of the observable states and animals who are in the unobservable state(s) at the same occasion? The path from capture-recapture probabilities to a multinomial mixture and bases structure was proven in Pradel et al. (2003) and we retain the same framework: this is illustrated using the toy-example from Table 5.1 (also presented in Chapter 4) as a starting point and setting the capture probability in state C to 0 (thus rendering C unobservable); Table 5.2 illustrates how this unobservable state translates into an unobservable basis. We use the following colour coding: red for the observable bases, blue for the mixtures and grey for the unobservable bases.

Testing for unobservable states means testing for unobservable bases in the more general mixture and bases framework, introduced in Chapter 4. Despite the apparent simplicity, the procedure was not a straightforward extension of Test M and many steps were needed. Hence, the work naturally shifted to focus on the more general model structure of mixtures of multinomials

Table 5.1: Table of the multi-state m-array terms associated to Test M at occasion $i = 2$. The mixtures are denoted by M and the bases by B.

m_{13}^{AA}	m_{13}^{AB}	m_{13}^{AC}	m_{14}^{AA}	m_{14}^{AB}	m_{14}^{AC}	M1
m_{13}^{BA}	m_{13}^{BB}	m_{13}^{BC}	m_{14}^{BA}	m_{14}^{BB}	m_{14}^{BC}	M2
m_{13}^{CA}	m_{13}^{CB}	m_{13}^{CC}	m_{14}^{CA}	m_{14}^{CB}	m_{14}^{CC}	M3
m_{23}^{AA}	m_{23}^{AB}	m_{23}^{AC}	m_{24}^{AA}	m_{24}^{AB}	m_{24}^{AC}	B1
m_{23}^{BA}	m_{23}^{BB}	m_{23}^{BC}	m_{24}^{BA}	m_{24}^{BB}	m_{24}^{BC}	B2
m_{23}^{CA}	m_{23}^{CB}	m_{23}^{CC}	m_{24}^{CA}	m_{24}^{CB}	m_{24}^{CC}	B3

Table 5.2: Table for mixtures and bases, at occasion $i = 2$, state C is unobservable. The mixtures are denoted by M and the bases by B.

m_{13}^{AA}	m_{13}^{AB}	m_{14}^{AA}	m_{14}^{AB}	M1
m_{13}^{BA}	m_{13}^{BB}	m_{14}^{BA}	m_{14}^{BB}	M2
m_{23}^{AA}	m_{23}^{AB}	m_{24}^{AA}	m_{24}^{AB}	B1
m_{23}^{BA}	m_{23}^{BB}	m_{24}^{BA}	m_{24}^{BB}	B2
m_{23}^{CA}	m_{23}^{CB}	m_{24}^{CA}	m_{24}^{CB}	B3

and bases, where samples from the mixtures and only some of the bases are available (no samples available for the unobservable bases). In order to do this, parameter redundancy was investigated. The concept is explained in Section 5.2 and general results are derived in Section 5.3. Using the parameter redundancy results, we go on to implement a testing procedure for one and two unobservable bases in Section 5.4. We assess whether it works as expected using simulations under very good conditions (i.e. large sample size). This procedure is then applied to a capture-recapture setting in Section 5.5, using simulation in the same way; we also apply the procedure to the Canada geese dataset. Finally we conclude and discuss our findings in Section 5.6. As shown by this chapter introduction, the primary focus shifted naturally as the work progressed. The chapter is mainly concerned with a procedure to test for one or two unobservable bases in a general mixture-bases context (although it could theoretically be used for more unobservable bases); and the procedure to test

for unobservable states in a capture-recapture setting becomes secondary as it constitutes an application of the more general results derived in this chapter.

5.2 Investigating parameter redundancy

5.2.1 Introduction

The cell probabilities associated with the frequency cells in Table 5.2 are given in Table 5.3. We use the MMLM approach (detailed in Chapter 4, Section 4.2) and first attempt to estimate the cell probabilities associated with the bases and the mixing probabilities via maximum-likelihood. Using a numerical example corresponding to the model presented in Table 5.3, the estimation of the 25 parameters (7×3 basis probabilities + 2×2 mixing probabilities) via maximum likelihood was found to be problematic. Indeed, different sets of initial values led to different parameter estimates for the same maximised likelihood value.

If different sets of maximum likelihood estimates (MLEs) result in the same optimum log-likelihood value, this means that there is a problem with identifiability. A model is globally identifiable if it is characterised by a unique set of MLEs: $M(\theta_1) = M(\theta_2)$ only if $\theta_1 = \theta_2$ (Cole et al., 2010). A non-identifiable model can be reparametrised using a smaller number of parameters (Cole et al., 2010). In other words, it is over-parametrised or parameter-redundant.

Table 5.3: Cell-probabilities for mixtures and bases with one unobservable basis

$\pi_1 p_1^{B1}+\pi_2 p_1^{B2}$ $+(1-\pi_1-\pi_2)p_1^{B3}$	$\pi_1 p_2^{B1}+\pi_2 p_2^{B2}$ $+(1-\pi_1-\pi_2)p_2^{B3}$	$\pi_1 p_3^{B1}+\pi_2 p_3^{B2}$ $+(1-\pi_1-\pi_2)p_3^{B3}$	$\pi_1(1- p_1^{B1}-p_2^{B1}-p_3^{B1})$ $+\pi_2(1-p_1^{B2}-p_2^{B2}-p_3^{B2})$ $+(1-\pi_1-\pi_2)(1-p_1^{B3}-p_2^{B3}-p_3^{B3})$	M1
$\gamma_1 p_1^{B1}+\gamma_2 p_1^{B2}$ $+(1-\gamma_1-\gamma_2)p_1^{B3}$	$\gamma_1 p_2^{B1}+\gamma_2 p_2^{B2}$ $+(1-\gamma_1-\gamma_2)p_2^{B3}$	$\gamma_1 p_3^{B1}+\gamma_2 p_3^{B2}$ $+(1-\gamma_1-\gamma_2)p_3^{B3}$	$\gamma_1(1-p_1^{B1}-p_2^{B1}-p_3^{B1})$ $+\gamma_2(1-p_1^{B2}-p_2^{B2}-p_3^{B2})$ $+(1-\gamma_1-\gamma_2)(1-p_1^{B3}-p_2^{B3}-p_3^{B3})$	M2
p_1^{B1}	p_2^{B1}	p_3^{B1}	$(1-p_1^{B1}-p_2^{B1}-p_3^{B1})$	B1
p_1^{B2}	p_2^{B2}	p_3^{B2}	$(1-p_1^{B2}-p_2^{B2}-p_3^{B2})$	B2
p_1^{B3}	p_2^{B3}	p_3^{B3}	$(1-p_1^{B3}-p_2^{B3}-p_3^{B3})$	B3

A parameter-redundant model has a flat ridge in the likelihood surface which results in multiple sets of MLEs (Catchpole and Morgan, 1997). Cole et al. (2010) showed that parameter redundancy within a model can be investigated by using an exhaustive summary $\boldsymbol{\kappa}$, which is a vector of parameters that defines the model uniquely. The exhaustive summary is differentiated with respect to the model parameters $\boldsymbol{\theta}$ to form the derivative matrix $\mathbf{D} = \partial\boldsymbol{\kappa}/\partial\boldsymbol{\theta}$. The rank of $\mathbf{D}$ determines whether the model is parameter redundant: if it is lower than the number of parameters p, the model is parameter redundant; if the rank is equal to the number of parameters, the model is termed full rank.

Recall that each row of the model examined in this chapter (see example Table 5.3) corresponds to a multinomial distribution (Yantis et al., 1991). An exhaustive summary is therefore a vector containing all the cell-probabilities, noting that the more complicated last column of each row (defined by subtracting the sum of row probabilities from 1) can be omitted since these terms do not affect the rank of the derivative matrix (Catchpole and Morgan, 1997). For instance, an exhaustive summary for our example from Table 5.3 would be:

$$\boldsymbol{\kappa}(\boldsymbol{\theta}) = \begin{bmatrix} \pi_1 p_1^{B1} + \pi_2 p_1^{B2} + (1-\pi_1-\pi_2) p_1^{B3} \\ \pi_1 p_2^{B1} + \pi_2 p_2^{B2} + (1-\pi_1-\pi_2) p_2^{B3} \\ \vdots \\ p_3^{B2} \\ (1 - p_1^{B2} - p_2^{B2} - p_3^{B2}) \end{bmatrix}$$

The derivative matrix also provides the independent parameter combinations that can be estimated. These allow the reformulation of a non-parameter-redundant model in terms of a smaller number of parameters. Cole et al. (2010) detail the steps of detecting parameter redundancy and finding the estimable parameter combinations, in their Theorem 1, *Testing parameter redundancy*, which is given below:

Theorem 1 (Testing parameter redundancy)

(a) *(i) If* $\boldsymbol{D}$ *has rank equal to* p *then the model is full rank.*

(ii) If the rank of $\boldsymbol{D}$ *is equal to* $q < p$*, then the model is parameter redundant. There are* q *estimable parameters and the model has deficiency* $d = p - q$*.*

(b) If the model is parameter redundant the estimable parameters can be determined by solving $\boldsymbol{\alpha}(\boldsymbol{\theta})^T \boldsymbol{D}(\boldsymbol{\theta}) = 0$*, which has* d *solutions, labelled* $\alpha_j(\boldsymbol{\theta})$ *for* $j = 1,\ldots, d$*, with individual entries* $\alpha_{ij}(\boldsymbol{\theta})$*. Any* $\alpha_{ij}(\boldsymbol{\theta})$ *which are zero for all* d *solutions correspond to a parameter,* θ_i*, which is estimable. The solutions of the system of linear first-order partial differential equations (PDEs),*

$$\sum_{i=1}^{p} \alpha_{ij} \frac{\partial f}{\partial \theta_i} = 0 \text{ } j\text{=}1,\ldots,d \text{ and } f \text{ an arbitrary function}$$

form the set of estimable parameters. Parameterised in terms of the estimable parameters, the model is full rank.

The symbolic algebra corresponding to these steps is carried out using the software Maple. We note that the parameter redundancy of multi-state capture-recapture models in the presence of unobservable states has been investigated by Cole (2012), and the paper provides estimable combinations of capture-recapture parameters. Our aim here is different, as we derive parameter redundancy results for the more general framework of mixtures of multinomials and their associated bases, when some of the bases are unobservable.

5.2.2 Parameter redundancy results for an example model: 4 bins, 2 mixtures, 3 bases - 2 observable and 1 unobservable

There are obvious cases when the model will be parameter-redundant because there is not enough information to estimate all the parameters uniquely. When all the bases are observable (i.e. usual product-multinomial context), if the exhaustive summary has fewer distinct terms than the number of parameters, the model will be parameter redundant (Cole et al., 2012). Let K be the number of columns (bins) of the multinomials, let B be the number of bases and M the number of mixtures. The number of distinct independent terms in the exhaustive summary is $(B+M)(K-1)$ whilst the number of parameters to be estimated is $B(K-1)$ basis probabilities and $M(B-1)$ mixing probabilities. This results in the inequality $(B+M)(K-1) \geq B(K-1)+M(B-1)$ which simplifies to $K \geq B$. Therefore when there are no unobservable bases, the number of bins needs to be at least equal to the number of bases in order to have enough information to estimate the parameters.

As a starting point to examine parameter redundancy, we use a model that could be identifiable when all bases are observable (i.e. $K \geq B$) and examine the case of $M = 2$, $B = 3$ formed of two observable bases, denoted by $O = 2$ and one unobservable basis, denoted by $U = 1$; with $K = 4$. This corresponds to the model structure presented in Table 5.3 and the set of model parameters is

$$\boldsymbol{\theta} = \begin{bmatrix} \gamma_1 & \gamma_2 & p_1^{B1} & p_2^{B1} & p_3^{B1} & p_1^{B2} & p_2^{B2} & p_3^{B2} & p_1^{B3} & p_2^{B3} & p_3^{B3} & \pi_1 & \pi_2 \end{bmatrix}^T.$$

Let $\boldsymbol{P}$ denote the matrix of cell-probabilities of the model, with

$$\boldsymbol{P} = \begin{bmatrix} \pi_1 p_1^{B1} + \pi_2 p_1^{B2} + (1-\pi_1-\pi_2) p_1^{B3} & \pi_1 p_2^{B1} + \pi_2 p_2^{B2} + (1-\pi_1-\pi_2) p_2^{B3} & \pi_1 p_3^{B1} + \pi_2 p_3^{B2} + (1-\pi_1-\pi_2) p_3^{B3}) & \pi_1 (1-p_1^{B1}-p_2^{B1}-p_3^{B1}) + \pi_2 (1-p_1^{B2}-p_2^{B2}-p_3^{B2}) + (1-\pi_1-\pi_2)(1-p_1^{B3}-p_2^{B3}-p_3^{B3}) \\ \gamma_1 p_1^{B1} + \gamma_2 p_1^{B2} + (1-\gamma_1-\gamma_2) p_1^{B3} & \gamma_1 p_2^{B1} + \gamma_2 p_2^{B2} + (1-\gamma_1-\gamma_2) p_2^{B3} & \gamma_1 p_3^{B1} + \gamma_2 p_3^{B2} + (1-\gamma_1-\gamma_2) p_3^{B3} & \gamma_1 (1-p_1^{B1}-p_2^{B1}-p_3^{B1}) + \gamma_2 (1-p_1^{B2}-p_2^{B2}-p_3^{B2}) + (1-\gamma_1-\gamma_2)(1-p_1^{B3}-p_2^{B3}-p_3^{B3}) \\ p_1^{B1} & p_2^{B1} & p_3^{B1} & (1-p_1^{B1}-p_2^{B1}-p_3^{B1}) \\ p_1^{B2} & p_2^{B2} & p_3^{B2} & (1-p_1^{B2}-p_2^{B2}-p_3^{B2}) \end{bmatrix}. \tag{5.1}$$

As noted previously, the vector denoted $\boldsymbol{\kappa}(\boldsymbol{\theta})$ containing all of the cell-probabilities constitutes an exhaustive summary for this model:

$$\boldsymbol{\kappa}(\boldsymbol{\theta}) = \begin{bmatrix} \pi_1 p_1^{B1} + \pi_2 p_1^{B2} + (1-\pi_1-\pi_2) p_1^{B3} \\ \pi_1 p_2^{B1} + \pi_2 p_2^{B2} + (1-\pi_1-\pi_2) p_2^{B3} \\ \vdots \\ p_3^{B2} \\ (1-p_1^{B2}-p_2^{B2}-p_3^{B2}) \end{bmatrix}$$

Using Maple to perform symbolic computations, the first order derivative matrix $\boldsymbol{D}$ is

$$
\boldsymbol{D} = \begin{bmatrix}
0 & 0 & 0 & 0 & p_1^{B1} - p_1^{B3} & p_2^{B1} - p_2^{B3} & p_3^{B1} - p_3^{B3} & -p_1^{B1} - p_2^{B1} - p_3^{B1} + p_1^{B3} + p_2^{B3} + p_3^{B3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p_1^{B2} - p_1^{B3} & p_2^{B2} - p_2^{B3} & p_3^{B2} - p_3^{B3} & -p_1^{B2} - p_2^{B2} - p_3^{B2} + p_1^{B3} + p_2^{B3} + p_3^{B3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\pi_1 & 0 & 0 & -\pi_1 & \gamma_1 & 0 & 0 & -\gamma_1 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & \pi_1 & 0 & -\pi_1 & 0 & \gamma_1 & 0 & -\gamma_1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & \pi_1 & -\pi_1 & 0 & 0 & \gamma_1 & -\gamma_1 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
\pi_2 & 0 & 0 & -\pi_2 & \gamma_2 & 0 & 0 & -\gamma_2 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\
0 & \pi_2 & 0 & -\pi_2 & 0 & \gamma_2 & 0 & -\gamma_2 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\
0 & 0 & \pi_2 & -\pi_2 & 0 & 0 & \gamma_2 & -\gamma_2 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\
1 - \pi_1 - \pi_2 & 0 & 0 & -1 + \pi_1 + \pi_2 & 1 - \gamma_1 - \gamma_2 & 0 & 0 & -1 + \gamma_1 + \gamma_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 - \pi_1 - \pi_2 & 0 & -1 + \pi_1 + \pi_2 & 0 & 1 - \gamma_1 - \gamma_2 & 0 & -1 + \gamma_1 + \gamma_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 - \pi_1 - \pi_2 & -1 + \pi_1 + \pi_2 & 0 & 0 & 1 - \gamma_1 - \gamma_2 & -1 + \gamma_1 + \gamma_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
p_1^{B1} - p_1^{B3} & p_2^{B1} - p_2^{B3} & p_3^{B1} - p_3^{B3} & -p_1^{B1} - p_2^{B1} - p_3^{B1} + p_1^{B3} + p_2^{B3} + p_3^{B3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
p_1^{B2} - p_1^{B3} & p_2^{B2} - p_2^{B3} & p_3^{B2} - p_3^{B3} & -p_1^{B2} - p_2^{B2} - p_3^{B2} + p_1^{B3} + p_2^{B3} + p_3^{B3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

The rank of the derivative matrix is $q = 11$ while there are $p = 13$ parameters to be estimated. Applying Theorem 1a, there is a deficiency of $d = 13 - 11 = 2$, the model is parameter redundant. Note that in this case ($K = 4$), parameter redundancy would be expected due to the number of cells since there are 12 independent cells but 13 parameters.

In order to find an expression of the estimable parameter combinations, as per Theorem 1b, we solve $\boldsymbol{\alpha}^T \boldsymbol{D} = 0$, which will have d solutions. The positions of the zeros common to these solutions indicate which parameters are estimable (Catchpole et al., 1998; Cole et al., 2010). In this case, the two solutions are:

$$\boldsymbol{\alpha}_1^T = \begin{bmatrix} 0 & \frac{-1+\gamma_1+\gamma_2}{-1+\pi_1+\pi 2} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{p_1^{B2}-p_1^{B3}}{-1+\pi_1+\pi_2} & \frac{p_2^{B2}-p_2^{B3}}{-1+\pi_1+\pi_2} & \frac{p_3^{B2}-p_3^{B3}}{-1+\pi_1+\pi_2} & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{\alpha}_2^T = \begin{bmatrix} \frac{-1+\gamma_1+\gamma_2}{-1+\pi_1+\pi 2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{p_1^{B1}-p_1^{B3}}{-1+\pi_1+\pi_2} & \frac{p_2^{B1}-p_2^{B3}}{-1+\pi_1+\pi_2} & \frac{p_3^{B1}-p_3^{B3}}{-1+\pi_1+\pi_2} & 1 & 0 \end{bmatrix}.$$

The zeroes in positions 3 to 8 for both solutions obtained indicate that we can estimate the parameters p_1^{B1}, p_2^{B1}, p_3^{B1}, p_1^{B2}, p_2^{B2}, p_3^{B2}. The cell-probabilities of the observable bases $B1$ and $B2$ being estimable separately is as expected since they are both independent multinomials with enough independent cells. In contrast, the mixing probabilities and the cell-probabilities of the unobservable basis B3 are not estimable separately. The estimable combinations of parameters are found by solving a system of differential equations as per Theorem 1b: $\sum_{i=1}^{p} \alpha_{ij} \partial f / \partial \theta_i = 0$ for $j = 1, \ldots, d$ (Catchpole et al., 1998; Cole et al., 2010). In this case, as the deficiency is $d = 2$, the two differential equations are:

$$\begin{cases} \frac{-1+\gamma_1+\gamma_2}{-1+\pi_1+\pi 2} \frac{\partial f}{\partial \gamma_2} + \frac{p_1^{B2}-p_1^{B3}}{-1+\pi_1+\pi_2} \frac{\partial f}{\partial p_1^{B3}} + \frac{p_2^{B2}-p_2^{B3}}{-1+\pi_1+\pi_2} \frac{\partial f}{\partial p_2^{B3}} + \frac{p_3^{B2}-p_3^{B3}}{-1+\pi_1+\pi_2} \frac{\partial f}{\partial p_3^{B3}} + \frac{\partial f}{\partial \pi_2} = 0 \\ \frac{-1+\gamma_1+\gamma_2}{-1+\pi_1+\pi 2} \frac{\partial f}{\partial \gamma_1} + \frac{p_1^{B1}-p_1^{B3}}{-1+\pi_1+\pi_2} \frac{\partial f}{\partial p_1^{B3}} + \frac{p_2^{B1}-p_2^{B3}}{-1+\pi_1+\pi_2} \frac{\partial f}{\partial p_2^{B3}} + \frac{p_3^{B1}-p_3^{B3}}{-1+\pi_1+\pi_2} \frac{\partial f}{\partial p_3^{B3}} + \frac{\partial f}{\partial \pi_1} = 0 \end{cases}$$

The solution gives us the following estimable combinations of the original parameters:

$$\begin{cases} \gamma_1(-p_1^{B1}+p_1^{B3})+\gamma_2(-p_1^{B2}+p_1^{B3})-p_1^{B3}(=-p_1^{M2}) \\ \gamma_1(-p_2^{B1}+p_2^{B3})+\gamma_2(-p_2^{B2}+p_2^{B3})-p_2^{B3}(=-p_2^{M2}) \\ \gamma_1(-p_3^{B1}+p_3^{B3})+\gamma_2(-p_3^{B2}+p_3^{B3})-p_3^{B3}(=-p_3^{M2}) \\ \frac{-1+\pi_1+\pi_2}{-1+\gamma_1+\gamma_2} \\ \frac{(1-\pi_1)\gamma_2+\pi_2(\gamma_1-1)}{-1+\gamma_1+\gamma_2} \end{cases}$$

The three first estimable combinations are easily recognizable as the cell-probabilities of mixture M2, denoted by p_i^{M2}. Let $\beta_1 = (-1+\pi_1+\pi_2)\,/\,(-1+\gamma_1+\gamma_2)$ and $\beta_2 = [(1-\pi_1)\gamma_2+\pi_2(\gamma_1-1)]\,/\,[-1+\gamma_1+\gamma_2]$, the cell-probabilities of mixture M1, denoted by p_i^{M1} can be expressed as a linear combination of the cell-probabilities of M2 and of the observable bases: $p_i^{M1} = \beta_1 p_i^{M2} + \beta_2 p_i^{B2} + (1-\beta_1-\beta_2)p_i^{B1}$. The cell-probabilities are needed for the likelihood construction so we reparametrise the model, creating a full rank parametrisation based on the estimable parameter combinations. A reparametrisation of $\boldsymbol{\theta}$ is given by $\boldsymbol{s}(\boldsymbol{\theta})$:

$$\boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} p_1^{B1} & p_1^{B2} & p_1^{M2} & p_2^{B1} & p_2^{B2} & p_2^{M2} & p_3^{B1} & p_3^{B2} & p_3^{M2} & \beta_1 & \beta_2 \end{bmatrix}^T$$

The exhaustive summary expressed in terms of this reparametrisation is:

$$\boldsymbol{\kappa}(\boldsymbol{s}) = \begin{bmatrix} \beta_1 p_1^{M2}+\beta_2 p_1^{B2}+(1-\beta_1-\beta_2)p_1^{B1} \\ \beta_1 p_2^{M2}+\beta_2 p_2^{B2}+(1-\beta_1-\beta_2)p_2^{B1} \\ \vdots \\ p_2^{B2} \\ p_3^{B2} \\ (1-p_1^{B2}-p_2^{B2}-p_3^{B2}) \end{bmatrix} \tag{5.2}$$

Parameter redundancy within this reparametrised model can be examined, using Cole et al. (2010)'s reparametrisation theorem, which is given below.

Theorem 2 (Reparametrisation theorem) *Let $\boldsymbol{D_s}$ denote the rank of the first-order derivative of the reparametrised exhaustive summary with respect to the reparametrisation $\boldsymbol{s}(\boldsymbol{\theta})$: $\boldsymbol{D_s} = \partial\boldsymbol{\kappa}(\boldsymbol{s})/\boldsymbol{s}$. Let $p_s = dim(\boldsymbol{s})$*

(a) (i) If the rank of D_s, r_s, is equal to p_s, $\boldsymbol{s}$ constitutes a reduced form exhaustive summary of the original model, and Theorem 1 may be applied to examine model structure.

(ii) If r_s is lower than p_s, $\boldsymbol{s}$ is not a reduced form exhaustive summary. A reduced-form exhaustive summary may be found by first solving $\boldsymbol{\alpha}^T\boldsymbol{D_s} = 0$ and then solving the appropriate partial differential equations as in Theorem 1b.

(b) If $rank(\partial\boldsymbol{s}/\partial\boldsymbol{\theta}) = p_s$, the number of estimable parameters is equal to r_s. If $r_s = dim(\boldsymbol{\theta})$then the model in terms of $\boldsymbol{\theta}$ is full rank. If $r_s < dim(\boldsymbol{\theta})$ the model in terms of $\boldsymbol{\theta}$ is parameter redundant.

Here, the rank of D_s is 11, for 11 parameters in the reparametrisation (p_s=11). By application of Theorem 2a, $\boldsymbol{s}$ constitutes a reduced-form exhaustive summary, and the reparametrised model is of full rank (by application of Theorem 1a).

5.3 Generalising the parameter redundancy results

Recalling the notations used earlier: let K denote the number of bins of the multinomials, M the number of mixtures, B the number of bases. We introduce additional notations to distinguish observable and unobservable bases: respectively denoted by O and U ($B = O + U$). In this section, we wish to obtain more general parameter redundancy results for a product multinomial model consisting of mixtures and associated bases, when some of the bases are

unobservable. The general model is presented in Table 5.4 in terms of cell-probabilities, recalling that the mixing probabilities and the row-probabilities both sum to one. We build on the example from Section 5.2.2, successively expanding the table in different directions, which leads to more general parameter redundancy results. Thus parameter redundancy is examined in Section 5.3.1 for an increase in the number of bins K, in Section 5.3.2 for an increase in the number of mixtures M, in Sections 5.3.3 and 5.3.4 for respectively an increase in the number of observable bases and unobservable bases. Finally, a summary of the results is collated in Section 5.3.5.

In order to generalise the parameter results, we use the reparametrisation theorem (Theorem 2) combined with Theorem 3, the *extension theorem*. The extension theorem applies to full rank models and is cited below from Cole et al. (2010).

Theorem 3 (Extension theorem) *If a model with parameters $\boldsymbol{\theta_1}$, exhaustive summary $\boldsymbol{\kappa}(\boldsymbol{\theta_1})$, and derivative matrix $\boldsymbol{D_1}(\boldsymbol{\theta_1}) = \partial\boldsymbol{\kappa_1}/\partial\boldsymbol{\theta_1}$, is extended, adding extra parameters $\boldsymbol{\theta_2}$, with the exhaustive summary extended to $\boldsymbol{\kappa}(\boldsymbol{\theta'}) = \left[\boldsymbol{\kappa_1}(\boldsymbol{\theta_1}), \boldsymbol{\kappa_2}(\boldsymbol{\theta'})\right]$, with $\boldsymbol{\theta'} = [\boldsymbol{\theta_1}, \boldsymbol{\theta_2}]$. The derivative matrix of the extended model is $\boldsymbol{D} = \begin{bmatrix} \boldsymbol{D_1}(\boldsymbol{\theta_1}) & \boldsymbol{D_{2,1}}(\boldsymbol{\theta_1}) \\ 0 & \boldsymbol{D_{2,2}}(\boldsymbol{\theta_2}) \end{bmatrix}$, with $\boldsymbol{D_{2,1}} = \frac{\partial\boldsymbol{\kappa_2}}{\partial\boldsymbol{\theta_1}}$ and $\boldsymbol{D_{2,2}} = \frac{\partial\boldsymbol{\kappa_2}}{\partial\boldsymbol{\theta_2}}$. If the original model is full rank (i.e. $\boldsymbol{D_1}$ is full rank), and $\boldsymbol{D_{2,2}}$ is full rank, then the extended model is full rank also.*

If the model examined is parameter redundant, we first reparametrise it so as to obtain a full rank model. The extension theorem can then be applied to the reduced-form exhaustive summary obtained from the reparametrisation theorem (Cole et al., 2010).

Table 5.4: Cell-probabilities for a general model with M mixtures, K bins, B bases amongst which O are observable and U are unobservable: $\pi_{O_j}^{M_m}$ denotes the mixing probability associated with observable basis O_j for mixture M_m, and similarly $\pi_{U_u}^{M_m}$ the mixing probability associated to unobservable basis U_u, $p_k^{O_j}$ and $p_k^{U_u}$ respectively denote the cell-probability of the kth bin of observable basis O_j and unobservable basis U_u

$\pi_{O_1}^{M_1}p_1^{O_1}+\ldots+\pi_{O_O}^{M_1}p_1^{O_O}+\pi_{U_1}^{M_1}p_1^{U_1}+\ldots+(1-\sum_{o=1}^{O}\pi_{O_o}^{M1}-\sum_{u=1}^{U-1}\pi_{U_u}^{M_1})p_1^{U_U}$	$\ldots$	$\pi_{O_1}^{M_1}(1-\sum_{k=1}^{K-1}p_k^{O_1})+\ldots+\pi_{O}^{M_1}(1-\sum_{k=1}^{K-1}p_k^{O_O})+\pi_{U_1}^{M_1}(1-\sum_{k=1}^{K-1}p_k^{U_1})+\ldots+(1-\sum_{o=1}^{O}\pi_{O_o}^{M1}-\sum_{u=1}^{U-1}\pi_{U_u}^{M_1})(1-\sum_{k=1}^{K-1}p_k^{U_U})$	M_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\pi_{O_1}^{M_M}p_1^{O_1}+\ldots+\pi_{O_O}^{M_M}p_1^{O_O}+\pi_{U_1}^{M_M}p_1^{U_1}+\ldots+(1-\sum_{o=1}^{O}\pi_{O_o}^{M_M}-\sum_{u=1}^{U-1}\pi_{U_u}^{M_M})p_1^{U_U}$	$\ldots$	$\pi_{O_1}^{M_M}(1-\sum_{k=1}^{K-1}p_k^{O_1})+\ldots+\pi_{O_O}^{M_M}(1-\sum_{k=1}^{K-1}p_k^{O_O})+\pi_{U_1}^{M_M}(1-\sum_{k=1}^{K-1}p_k^{U_1})+\ldots+(1-\sum_{o=1}^{O}\pi_{O_o}^{M_M}-\sum_{u=1}^{U-1}\pi_{U_u}^{M_M})(1-\sum_{k=1}^{K-1}p_k^{U_U})$	M_M
$p_1^{O_1}$	$\ldots$	$(1-\sum_{k=1}^{K-1}p_k^{O_1})$	O_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$p_1^{O_O}$	$\ldots$	$(1-\sum_{k=1}^{K-1}p_k^{O_O})$	O_O
$p_1^{U_1}$	$\ldots$	$(1-\sum_{k=1}^{K-1}p_k^{U_1})$	U_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$p_1^{U_U}$	$\ldots$	$(1-\sum_{k=1}^{K-1}p_k^{U_U})$	U_U

5.3.1 Increasing the number of bins

In Section 5.2.2, we have shown that for $K = 4$, a multinomial model structure comprising of $M = 2$, $O = 2$, $U = 1$ presents a deficiency of 2. Hence, we start with Equation 5.2, the reparametrised exhaustive summary derived in Section 5.2.2 for $K = 4$. We present it in a slightly different format: first, we omit the terms from the last column from the model cell-probabilities $\boldsymbol{P}$ (Equation 5.1) since these terms are more complicated (one minus the row probabilities) and do not affect the rank of the derivative matrix as mentioned in Section 5.2 (Catchpole and Morgan, 1997). Then, we slightly modify the notations used before in order to distinguish observable and unobservable bases more clearly: the observable bases denoted B_1 and B_2 in the original Equation 5.1 become O_1 and O_2 whilst the unobservable basis B_3 is now denoted by U_1. This gives $\boldsymbol{\kappa_1}(\boldsymbol{s_1})$ as

$$\boldsymbol{\kappa_1}(\boldsymbol{s_1}) = \begin{bmatrix} \beta_1 p_1^{M2} + \beta_2 p_1^{O2} + (1-\beta_1-\beta_2) p_1^{O1} \\ \beta_1 p_2^{M2} + \beta_2 p_2^{O2} + (1-\beta_1-\beta_2) p_2^{O1} \\ \beta_1 p_3^{M2} + \beta_2 p_3^{O2} + (1-\beta_1-\beta_2) p_3^{O1} \\ p_1^{M2} \\ p_2^{M2} \\ p_3^{M2} \\ p_1^{O1} \\ p_2^{O1} \\ p_3^{O1} \\ p_1^{O2} \\ p_2^{O2} \\ p_3^{O2} \end{bmatrix},$$

where $\boldsymbol{s_1} = \begin{bmatrix} p_1^{M2} & p_2^{M2} & p_3^{M2} & p_1^{O1} & p_2^{O1} & p_3^{O1} & p_1^{O2} & p_2^{O2} & p_3^{O2} & \beta_1 & \beta_2 \end{bmatrix}^T$. When K is increased by 1 (i.e.Table 5.3 extended by one column), the additional cell-probabilities are:

$$\boldsymbol{\kappa_2}(\boldsymbol{\theta_2}) \begin{bmatrix} \pi_1 p_4^{O1} + \pi_2 p_4^{O2} + (1-\pi_1-\pi_2) p_4^{U1} \\ \gamma_1 p_4^{O1} + \gamma_2 p_4^{O2} + (1-\gamma_1-\gamma_2) p_4^{U1} \\ p_4^{O1} \\ p_4^{O2} \end{bmatrix}, \text{ where } \boldsymbol{\theta_2} = \begin{bmatrix} p_4^{U1} & p_4^{O1} & p_4^{O2} \end{bmatrix}^T.$$

Using the reparametrisation from Section 5.2.2, the new additional terms can be expressed as

$$\boldsymbol{\kappa_2}(\boldsymbol{s_2}) = \begin{bmatrix} \beta_1 p_4^{M2} + \beta_2 p_4^{O2} + (1-\beta_1-\beta_2) p_4^{O1} \\ p_4^{M2} \\ p_4^{O1} \\ p_4^{O2} \end{bmatrix},$$

and the additional parameters are $\boldsymbol{s_2} = \begin{bmatrix} p_4^{M2} & p_4^{O1} & p_4^{O2} \end{bmatrix}^T$. Recall from Section 5.2.2 that the derivative matrix of the initial reparametrised model $\boldsymbol{D_1}(\boldsymbol{s_1}) = \partial \boldsymbol{\kappa_1}(\boldsymbol{s_1})/\partial \boldsymbol{s_1}$ is of full rank. Let us study the rank of $\boldsymbol{D_2}(\boldsymbol{s_2})$:

$$\boldsymbol{D_2}(\boldsymbol{s_2}) = \frac{\partial \boldsymbol{\kappa_2}}{\partial \boldsymbol{s_2}} = \begin{bmatrix} \beta_1 & 1 & 0 & 0 \\ -\beta_1 - \beta_2 + 1 & 0 & 1 & 0 \\ \beta_2 & 0 & 0 & 1 \end{bmatrix}$$

The rank of this matrix is 3, which is equal to the number of additional parameters: $\boldsymbol{D_2}(\boldsymbol{s_2})$ is full rank. Note that $\boldsymbol{D_2}(\boldsymbol{s_2})$ corresponds to the term $\boldsymbol{D_{2,2}}$ from the extension theorem and $\boldsymbol{D_1}(\boldsymbol{s_1})$ to $\boldsymbol{D_1}$. Both these terms have been shown to be full rank, hence the reparametrised extended model is of full rank. It has three additional estimable parameters (p_4^{O1}, p_4^{O2}, p_4^{M2}) for three additional original parameters (p_4^{O1}, p_4^{O2}, p_4^{U1}). Since the extended reparametrised model is full rank, thus with rank 14, the extended model in terms of the 16 original parameters is parameter redundant with deficiency 2. By induction, the generalisation from a multinomial with K bins to $K+1$ bins will proceed in the same way (Cole et al., 2012). Hence, the extended model with $M = 2$, $O = 2$, $U = 1$, will always be parameter redundant with deficiency 2 for all

$K \geq 4$.

5.3.2 Increasing the number of mixtures

In this section, we aim to generalise the parameter redundancy results when increasing the number of mixtures. Starting from $M = 1$, $O = 2$, $U = 1$, $K = 4$, and labelling this mixture M_2 for convenience, the exhaustive summary can be expressed as

$$\boldsymbol{\kappa_{1M}}(\boldsymbol{\theta_{1M}}) = \begin{bmatrix} \gamma_1 p_1^{O1} + \gamma_2 p_1^{O2} + (1-\gamma_1-\gamma_2) p_1^{U1} \\ \gamma_1 p_2^{O1} + \gamma_2 p_2^{O2} + (1-\gamma_1-\gamma_2) p_2^{U1} \\ \gamma_1 p_3^{O1} + \gamma_2 p_3^{O2} + (1-\gamma_1-\gamma_2) p_3^{U1} \\ p_1^{O1} \\ p_2^{O1} \\ p_3^{O1} \\ p_1^{O2} \\ p_2^{O2} \\ p_3^{O2} \end{bmatrix}$$

with $\boldsymbol{\theta_{1M}} = \begin{bmatrix} \gamma_1 & \gamma_2 & p_1^{O1} & p_1^{O2} & p_1^{U1} & p_2^{O1} & p_2^{O2} & p_2^{U1} & p_3^{O1} & p_3^{O2} & p_3^{U1} \end{bmatrix}^T$. Using Maple, the derivative matrix of the exhaustive summary with respect to the parameters $\boldsymbol{D_{1M}} = \partial \boldsymbol{\kappa_{1M}}(\boldsymbol{\theta_{1M}})/\partial \boldsymbol{\theta_{1M}}$ has rank 9, it is parameter redundant with deficiency 2. Solving the appropriate partial differential equations, the estimable parameter combinations are again found to be the cell-probabilities of the mixture as in Section 5.2.2. We denote these estimable parameters by p_i^{M2}, with $p_i^{M2} = \gamma_1 p_i^{O1} + \gamma_2 p_i^{O2} + (1-\gamma_1-\gamma_2) p_i^{U1}$. The reparametrised exhaustive summary is then $\boldsymbol{\kappa_{1M}}(\boldsymbol{s_{1M}}) = \begin{bmatrix} p_1^{M2} & p_2^{M2} & p_3^{M2} & p_1^{O1} & p_2^{O1} & p_3^{O1} & p_1^{O2} & p_2^{O2} & p_3^{O2} \end{bmatrix}^T$, where $\boldsymbol{s_{1M}} = \begin{bmatrix} p_1^{M2} & p_2^{M2} & p_3^{M2} & p_1^{O1} & p_2^{O1} & p_3^{O1} & p_1^{O2} & p_2^{O2} & p_3^{O2} \end{bmatrix}^T$. In these reparametrised terms, the model simply becomes a product-multinomial with each cell-probability corresponding to a unique parameter: for instance multinomial M_2 is characterised solely by the p_i^{M2}, the same applies for multinomials

O_1 and O_2. This model, and hence $\boldsymbol{D_{1M}(s_{1M})} = \partial\boldsymbol{\kappa_{1M}(s_{1M})}/\partial\boldsymbol{s_{1M}}$ is therefore obviously full rank (Cole et al., 2012). Note that this result is consistent with the rank of the derivative matrix found using Maple. The extension theorem is now applied to this full-rank reparametrised model. When increasing the number of mixtures to 2, the extended summary in reparametrised terms is $\boldsymbol{\kappa_1(s_1)}$ from Section 5.3.1. The additional terms are denoted by

$$\boldsymbol{\kappa_{2M}(s_{2M})} = \begin{bmatrix} p_1^{M2}\beta_1 + \beta_2 p_1^{O2} + (1-\beta_1-\beta_2)p_1^{O1} \\ p_2^{M2}\beta_1 + \beta_2 p_2^{O2} + (1-\beta_1-\beta_2)p_2^{O1} \\ p_3^{M2}\beta_1 + \beta_2 p_3^{O2} + (1-\beta_1-\beta_2)p_3^{O1} \end{bmatrix}, \text{ where } \boldsymbol{s_{2M}} = \begin{bmatrix} \beta_1 & \beta_2 \end{bmatrix}^T.$$

Recall from Section 5.2.2 that $\beta_1 = (-1+\pi_1+\pi_2)/(-1+\gamma_1+\gamma_2)$ and $\beta_2 = [(1-\pi_1)\gamma_2 + \pi_2(\gamma_1-1)] \,/\, [-1+\gamma_1+\gamma_2]$. The derivative matrix term $\boldsymbol{D_{2,2}}$ from the extension theorem is

$$\boldsymbol{D_{2M}(s_{2M})} = \frac{\partial\boldsymbol{\kappa_{2M}}}{\partial\boldsymbol{s_{2M}}} = \begin{bmatrix} p_1^{M2} - p_1^{O1} & p_2^{M2} - p_2^{O1} & p_3^{M2} - p_3^{O1} \\ p_1^{O2} - p_1^{O1} & p_2^{O2} - p_2^{O1} & p_3^{O2} - p_3^{O1} \end{bmatrix}$$

The rank of this matrix is 2, which is equal to the number of additional parameters, thus $\boldsymbol{D_{2M}(s_{2M})}$ is full rank. Using the extension theorem, since $\boldsymbol{D_{1M}(s_{1M})}$ and $\boldsymbol{D_{2M}(s_{2M})}$ are both full rank, the extended reparametrised model is of full rank. It has two additional estimable parameters (β_1, β_2) for two additional original parameters (π_1, π_2). Since the extended reparametrised model is full rank, thus with rank 11, the extended model in terms of the 13 original parameters is parameter redundant with deficiency 2. By induction, the generalisation from a model with M mixtures to $M+1$ mixtures will proceed in the same way (Cole et al., 2012).

From the two applications of the extension theorem, we can now conclude that, a model with three bases, amongst which one is unobservable, with $K \geq 4$ bins and $M \geq 1$ mixtures will always be parameter redundant with deficiency 2.

5.3.3 Increasing the number of observable bases

We now increase the number of observable bases O: let $K = 4$, $U = 1$, $O = 3$, $M = 2$. The exhaustive summary $\boldsymbol{\kappa_{3O}}(\boldsymbol{\theta_{3O}})$ is obtained based on the general multinomial model presented in Table 5.4, once again omitting the last column.

$$\boldsymbol{\kappa_{3O}}(\boldsymbol{\theta_{3O}}) = \begin{bmatrix} \pi_1 p_1^{O1} + \pi_2 p_1^{O2} + \pi_3 p_1^{O3} + (1 - \pi_1 - \pi_2 - \pi_3) p_1^{U1} \\ \pi_1 p_2^{O1} + \pi_2 p_2^{O2} + \pi_3 p_2^{O3} + (1 - \pi_1 - \pi_2 - \pi_3) p_2^{U1} \\ \vdots \\ p_2^{U1} \\ p_3^{U1} \end{bmatrix}$$

with model parameters $\boldsymbol{\theta_{3O}} = \begin{bmatrix} \gamma_1 & \gamma_2 & \gamma_3 & p_1^{O1} & p_2^{O1} & \dots & p_2^{U1} & p_3^{U1} & \pi_1 & \pi_2 & \pi_3 \end{bmatrix}^T$.
The rank of the derivative matrix, $\boldsymbol{D_{3O}} = \partial \boldsymbol{\kappa_{3O}}(\boldsymbol{\theta_{3O}}) / \partial \boldsymbol{\theta_{3O}}$, is found to be 15 using Maple, whereas there are 18 parameters. The model is therefore parameter redundant with deficiency 3. Solving the appropriate set of partial differential equations as per Theorem 1b, the estimable parameter combinations can be expressed as:

$$\begin{cases} -p_1^{M2} = -\gamma_1 p_1^{O1} - \gamma_2 p_1^{O2} - \gamma_3 p_1^{O3} + (-1 + \gamma_1 + \gamma_2 + \gamma_3) p_1^{U1} \\ -p_2^{M2} = -\gamma_1 p_2^{O1} - \gamma_2 p_2^{O2} - \gamma_3 p_2^{O3} + (-1 + \gamma_1 + \gamma_2 + \gamma_3) p_2^{U1} \\ -p_3^{M2} = -\gamma_1 p_3^{O1} - \gamma_2 p_3^{O2} - \gamma_3 p_3^{O3} + (-1 + \gamma_1 + \gamma_2 + \gamma_3) p_3^{U1} \\ \beta_1 = \frac{-1 + \pi_1 + \pi_2 + \pi_3}{-1 + \gamma_1 + \gamma_2 + \gamma_3} \\ \beta_2 = \frac{(1 - \pi_1 - \pi_3)\gamma_2 + \pi_2(\gamma_1 + \gamma_3 - 1)}{-1 + \gamma_1 + \gamma_2 + \gamma_3} \\ \beta_3 = \frac{(1 - \pi_1 - \pi_2)\gamma_3 + \pi_3(\gamma_1 + \gamma_2 - 1)}{-1 + \gamma_1 + \gamma_2 + \gamma_3} \end{cases}$$

Once again, the cell-probabilities of mixture M_1 can be expressed as a linear combination of the cell-probabilities of M_2 and of the observable bases, with the βs being the coefficients:

$p_i^{M1} = p_i^{M2}\beta_1 + \beta_2 p_i^{O2} + \beta_3 p_i^{O3} + (1 - \beta_1 - \beta_2 - \beta_3) p_i^{O1}$.

A reparametrised exhaustive summary based on these estimable combinations can then expressed be expressed as:

$$\boldsymbol{\kappa_{3O}}(\boldsymbol{s_{3O}}) = \begin{bmatrix} p_1^{M2}\beta_1 + \beta_2 p_1^{O2} + \beta_3 p_1^{O3} + (1-\beta_1-\beta_2-\beta_3)p_1^{O1} \\ p_2^{M2}\beta_1 + \beta_2 p_2^{O2} + \beta_3 p_2^{O3} + (1-\beta_1-\beta_2-\beta_3)p_2^{O1} \\ p_3^{M2}\beta_1 + \beta_2 p_3^{O2} + \beta_3 p_3^{O3} + (1-\beta_1-\beta_2-\beta_3)p_3^{O1} \\ p_1^{M2} \\ p_2^{M2} \\ p_3^{M2} \\ p_1^{O1} \\ p_2^{O1} \\ p_3^{O1} \\ p_1^{O2} \\ p_2^{O2} \\ p_3^{O2} \\ p_1^{O3} \\ p_2^{O3} \\ p_3^{O3} \end{bmatrix},$$

with

$$\boldsymbol{s_{3O}} = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & p_1^{O1} & p_2^{O1} & p_3^{O1} & p_1^{O2} & p_2^{O2} & p_3^{O2} & p_1^{O3} & p_2^{O3} & p_3^{O3} & p_1^{M2} & p_2^{M2} & p_3^{M2} \end{bmatrix}^T.$$

The derivative matrix $\boldsymbol{D(s_{3O})} = \partial\boldsymbol{\kappa_{3O}}(\boldsymbol{s_{3O}})/\partial\boldsymbol{s_{3O}}$ is found to have rank 15. Hence, this reparametrised model is full rank as expected. Comparing the reparametrised exhaustive summary to the starting point $\boldsymbol{\kappa_1}(\boldsymbol{s_1})$ for the model with $O = 2$, $K = 4$, $U = 1$, $M = 2$, this is not a case where additional parameters are simply being added to the original exhaustive summary. Rather, the original exhaustive summary terms are also modified when increasing the number of observable bases. Hence, this situation does not lend itself to the application of the extension theorem.

However, we observe that the reparametrised exhaustive summary $\boldsymbol{\kappa_1}(\boldsymbol{s_1})$, which corresponds to a model with 2 observable bases, may be obtained from $\boldsymbol{\kappa_{3O}}(\boldsymbol{s_{3O}})$, by setting the following constraints: $\beta_3 = 0$ and $p_i^{O3} = 0$, for $i = 1, ..., 3$. The rank of the derivative matrix $\boldsymbol{D(s_{3O})}$ evaluated for the

constrained case would yield a rank of $15 - 4 = 11$ which is indeed the rank of $\boldsymbol{D_1}(\boldsymbol{s_1})$ (see Section 5.3.1).

By trying out situations with an increase in the number of observable bases to 4, 5 etc. as well as in the number of mixtures and bins, keeping U1 as the unobservable basis with associated mixing probability equal to one minus the others, we observe a pattern for the estimable combinations. For O observable bases, one unobservable basis ($U = 1$), M mixtures and K bins, these can be expressed using the general notation defined in Table 5.4, as

- the first $(K - 1)$ cell-probabilities of one of the mixtures M_l: p_i^{Ml} for $i = 1, ..., K - 1$ (this can be any mixture)
- and if $M \geq 2$, $(M - 1)(B - 1)$ parameters, with each of the remaining mixtures $M_{m;(m\neq l)}$ expressed as a linear combination of mixture M_l and the observable bases, using the following coefficients:
  - the coefficient of mixture M_l,

$$\beta_1^{Mm;m\neq l} = \frac{1 - \sum_{k=1}^{O} \pi_{Ok}^{Mm}}{1 - \sum_{k=1}^{O} \pi_{Ok}^{Ml}},$$

    (recall that π_{Ok}^{Mm} denotes the mixing probability of basis O_k for mixture M_m)
  - the coefficients of the observable bases O_j, for $j = 2, \ldots, O$

$$\beta_{Oj}^{Mm;m\neq l} = \frac{\pi_{Oj}^{Mm}\left(-1 + \sum_{\substack{k=1\\k\neq j}}^{O} \pi_{Ok}^{Ml}\right) + \pi_{Oj}^{Ml}\left(1 - \sum_{\substack{k=1\\k\neq j}}^{O} \pi_{Ok}^{Mm}\right)}{-(1 - \sum_{k=1}^{O} \pi_{Ok}^{Ml})}$$

  - the coefficient of observable basis O_1 is then expressed as

$$\beta_{O1}^{Mm;m\neq l} = 1 - \beta_1^{Mm;m\neq l} - \sum_{j=2}^{O} \beta_{Oj}^{Mm;m\neq l}$$

We use induction to find a general result regarding the deficiency of the

model with O observable bases. We have shown that a model with $M = 2$, $K = 4$, $O = 2$, $U = 1$ presents a deficiency of 2, and that a model with $M = 2$, $K = 4$, $O = 3$, $U = 1$ presents deficiency of 3. Suppose that a model with M mixtures and K bins, with O observable bases and one unobservable basis, presents a deficiency of $O + U - 1 = B - 1$, with B being the total number of bases in the model. For the induction step, we show that reducing the number of observable bases from O to $O - 1$ and thus the total number of bases from B to $B - 1$ whilst keeping all other dimensions fixed, results in a deficiency of $O - 1 + U - 1 = B - 2$. Keeping the other dimensions fixed, the reduced form exhaustive summary for a model with $O - 1$ observable bases can be obtained by constraining to 0 the $K - 1$ cell probabilities of basis O_O as well as the $(M-1)$ $\beta_O^{m \neq l}$ coefficients relating to basis O_O (for the $(M-1)$ mixtures expressed as linear combinations of the observable bases and mixture M_l in the reparametrised model). The rank of the derivative matrix of the exhaustive summary can be expressed as $r = p - d$, with p the number of parameters and d the deficiency. Therefore, the rank of the derivative matrix of the exhaustive summary for the model with O observable bases, with respect to the original parameters, will be $r = [M(B-1) + (K-1)B] - (B-1)$. Constraining the parameters pertaining to basis O to 0 ($K - 1$ cell probabilities and $M - 1$ coefficients), it follows that the rank of the derivative matrix model with $O-1$ observable bases is $r_2 = [M(B-1)+(K-1)B]-(B-1)-(K-1)-(M-1)$. The number of parameters to estimate for a model with $O-1$ observable bases and 1 unobservable basis (i.e. $B-1$ bases in total) is equal to $M(B-2)+(B-1)(K-1)$. Therefore, using the relationship $d = p - r$, the deficiency of the model with $O - 1$ observable bases is equal to $d = (M(B-2) + (B-1)(K-1)) - (M(B-1) + (K-1)B - (B-1) - (K-1) - (M-1)) = B - 2$. Hence a model with $B - 1$ bases amongst which one is unobservable ($O = B - 2$, $U = 1$) presents a deficiency of $B - 2$. Thus, we have shown by induction that a model with M mixtures, K bins, B bases, amongst which 1 is unobservable

$(O = B - 1, U = 1)$, is parameter redundant with deficiency $B - 1$.

5.3.4 Increasing the number of unobservable bases

Finally, we examine the effect of rendering another basis unobservable: $O = 2$, $U = 2$, $K = 4$, $M = 2$, again using the *testing parameter redundancy* theorem. The exhaustive summary for this model is:

$$\boldsymbol{\kappa_{2U}}(\boldsymbol{\theta_{2U}}) = \begin{bmatrix} \pi_1 p_1^{O1} + \pi_2 p_1^{O2} + \pi_3 p_1^{U1} + (1 - \pi_1 - \pi_2 - \pi_3) p_1^{U2} \\ \pi_1 p_2^{O1} + \pi_2 p_2^{O2} + \pi_3 p_2^{U1} + (1 - \pi_1 - \pi_2 - \pi_3) p_2^{U2} \\ \pi_1 p_3^{O1} + \pi_2 p_3^{O2} + \pi_3 p_3^{U1} + (1 - \pi_1 - \pi_2 - \pi_3) p_3^{U2} \\ \gamma_1 p_1^{O1} + \gamma_2 p_1^{O2} + \gamma_3 p_1^{U1} + (1 - \gamma_1 - \gamma_2 - \gamma_3) p_1^{U2} \\ \gamma_1 p_2^{O1} + \gamma_2 p_2^{O2} + \gamma_3 p_2^{U1} + (1 - \gamma_1 - \gamma_2 - \gamma_3) p_2^{U2} \\ \gamma_1 p_3^{O1} + \gamma_2 p_3^{O2} + \gamma_3 p_3^{U1} + (1 - \gamma_1 - \gamma_2 - \gamma_3) p_3^{U2} \\ p_1^{O1} \\ p_2^{O1} \\ p_3^{O1} \\ p_1^{O2} \\ p_2^{O2} \\ p_3^{O2} \end{bmatrix}$$

with the model parameters being

$$\boldsymbol{\theta_{2U}} = \begin{bmatrix} \gamma_1 & \gamma_2 & \gamma_3 & p_1^{O1} & p_2^{O1} & \dots & p_3^{U1} & p_1^{U2} & p_2^{U2} & p_3^{U2} & \pi_1 & \pi_2 & \pi_3 \end{bmatrix}^T$$

Using maple, the rank of the derivative matrix $\boldsymbol{D}_{2U}(\boldsymbol{\theta_{2U}}) = \partial \boldsymbol{\kappa_{2U}}(\boldsymbol{\theta_{2U}})/\partial \boldsymbol{\theta_{2U}}$ is 12 whilst there are 18 parameters to estimate, the model presents a deficiency of $d = 6$. When attempting to solve the partial differential equations necessary to obtain the estimable parameters, Maple runs out of memory. Therefore an incremental approach is used to find these estimable combinations. We refer back to a simpler model with the same total number of bases, amongst which only one is unobservable: $O = 3$, $U = 1$, $K = 4$, $M = 2$. We found a reduced-form exhaustive summary for this model in Section 5.3.3: $\boldsymbol{s_{3O}}$.

Keeping the total number of bases constant (here $B = 4$), if another basis becomes unobservable ($O = 2$ and $U = 2$), the number of parameters to estimate for the model won't change. However, the cell-probabilities associated to the newly unobservable basis will disappear from the exhaustive summary since they are no longer observed. Therefore, based on $\boldsymbol{\kappa_{3O}}(\boldsymbol{s_{3O}})$, we can express the model with 2 unobservable bases in terms of the reparametrised terms from $\boldsymbol{s_{3O}}$.

$$\boldsymbol{\kappa_{2U}}(\boldsymbol{s_{2U}}) = \begin{bmatrix} p_1^{M2}\beta_1 + \beta_2 p_1^{O2} + \beta_3 p_1^{U2} + (1-\beta_1-\beta_2-\beta_3)p_1^{O1} \\ p_2^{M2}\beta_1 + \beta_2 p_2^{O2} + \beta_3 p_2^{U2} + (1-\beta_1-\beta_2-\beta_3)p_2^{O1} \\ p_3^{M2}\beta_1 + \beta_2 p_3^{O2} + \beta_3 p_3^{U2} + (1-\beta_1-\beta_2-\beta_3)p_3^{O1} \\ p_1^{M2} \\ p_2^{M2} \\ p_3^{M2} \\ p_1^{O1} \\ p_2^{O1} \\ p_3^{O1} \\ p_1^{O2} \\ p_2^{O2} \\ p_3^{O2} \end{bmatrix}$$

with

$$\boldsymbol{s_{2U}} = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & p_1^{O1} & p_2^{O1} & p_3^{O1} & p_1^{O2} & p_2^{O2} & p_3^{O2} & p_1^{M2} & p_2^{M2} & p_3^{M2} & p_1^{U2} & p_2^{U2} & p_3^{U2} \end{bmatrix}^T.$$

Note that $\boldsymbol{s_{2U}}$ corresponds $\boldsymbol{s_{3O}}$, only with basis O_3 relabelled as U_2 since it is now unobservable.

The rank of the derivative matrix $\boldsymbol{D}_{2U}(\boldsymbol{s_{2U}}) = \partial \boldsymbol{\kappa_{2U}}(\boldsymbol{s_{2U}})/\partial \boldsymbol{s_{2U}}$ is 12, as expected, for 15 terms to estimate in $\boldsymbol{s_{2U}}$; $\boldsymbol{s_{2U}}$ is not a reduced form summary. Thus, applying Theorem 2a(ii), we solve the appropriate set of partial differential equations and find the following estimable parameter combinations:

$$\begin{cases} \tau_1 = p_1^{M2}\beta_1 + \beta_2 p_1^{O2} + \beta_3 p_1^{U2} + (-\beta_1 - \beta_2 - \beta_3)p_1^{O1} = p_1^{M1} - p_1^{O1} \\ \tau_2 = p_2^{M2}\beta_1 + \beta_2 p_2^{O2} + \beta_3 p_2^{U2} + (-\beta_1 - \beta_2 - \beta_3)p_2^{O1} = p_2^{M1} - p_2^{O1} \\ \tau_3 = p_3^{M2}\beta_1 + \beta_2 p_3^{O2} + \beta_3 p_3^{U2} + (-\beta_1 - \beta_2 - \beta_3)p_3^{O1} = p_3^{M1} - p_3^{O1} \end{cases}$$

A new reparametrisation in terms of these estimable combinations is given by

$$\boldsymbol{\kappa'_{2U}}(\boldsymbol{s'_{2U}}) = \begin{bmatrix} \tau_1 + p_1^{O1} & \tau_2 + p_2^{O1} & \tau_3 + p_3^{O1} & p_1^{M2} & p_2^{M2} & p_3^{M2} & p_1^{O1} & p_2^{O1} & p_3^{O1} & p_1^{O2} & p_2^{O2} & p_3^{O2} \end{bmatrix}^T$$

with

$$\boldsymbol{s'_{2U}} = \begin{bmatrix} \tau_1 & \tau_2 & \tau_3 & p_1^{M2} & p_2^{M2} & p_3^{M2} & p_1^{O1} & p_2^{O1} & p_3^{O1} & p_1^{O2} & p_2^{O2} & p_3^{O2} \end{bmatrix}^T.$$

The rank of the derivative matrix using this additional reparametrisation is 12, the newly reparametrised model is full rank and $\boldsymbol{s'_{2U}}$ is a reduced-form exhaustive summary for the model with 2 unobservable bases. The exhaustive summary for one unobservable basis $\boldsymbol{\kappa_{3O}}(\boldsymbol{s_{3O}})$ can now be expressed using the full-rank reparametrisation $\boldsymbol{s'_{2U}}$:

$$\boldsymbol{\kappa_{3O}}(\boldsymbol{s'_{3O}}) = \begin{bmatrix} \tau_1 + p_1^{O1} \\ \tau_2 + p_2^{O1} \\ \tau_3 + p_2^{O1} \\ p_1^{M2} \\ p_2^{M2} \\ p_3^{M2} \\ p_1^{O1} \\ p_2^{O1} \\ p_3^{O1} \\ p_1^{O2} \\ p_2^{O2} \\ p_3^{O2} \\ p_1^{O3} \\ p_2^{O3} \\ p_3^{O3} \end{bmatrix}$$

Hence, going from two unobservable bases to one unobservable basis, requires

the additional terms in the exhaustive summary $\boldsymbol{\kappa_2}(\boldsymbol{s'_{2,3O}}) = \begin{bmatrix} p_1^{O3} & p_2^{O3} & p_3^{O3} \end{bmatrix}^T$ for additional parameters to estimate $\boldsymbol{s'_{2,3O}} = \begin{bmatrix} p_1^{O3} & p_2^{O3} & p_3^{O3} \end{bmatrix}^T$, $\boldsymbol{\kappa_2}(\boldsymbol{s'_{2,3O}})$ is trivially of full rank. Therefore the extended model using this reparametrisation is of full rank when $U = 1$, $O = 3$. This means the number of estimable parameters is 15 for 18 original parameters, the original model with $U = 1$, $O = 3$ (i.e. in terms of $\boldsymbol{\theta_{3O}}$) is parameter redundant with deficiency 3, which is consistent with our findings from Section 5.3.3.

By induction, decreasing the number of unobservable bases from U to $U-1$ for a model with 4 bases, decreases the deficiency by 3 or, in other words, increasing the number of unobservable bases by one increases the deficiency by 3.

From the previous results in Section 5.3.3, a model with B bases amongst which one is unobservable yields a deficiency of $B - 1$. It follows that the deficiency increases by $B - 1$ when U increases by 1. Therefore, a model with B bases amongst which O are observable and U are unobservable will present a deficiency of $d = U(B - 1)$.

From experimenting with various example values of O and U, we deduced that the estimable combinations can be derived and expressed in a recursive manner. Suppose there are U unobservable bases within a total number of B bases, for O observable and M mixtures, we start from a model with 1 unobservable basis and $O + U - 1$ observable bases, for which it is easy to find a full-rank reparametrisation using the results from Section 5.3.3 (see for example $\boldsymbol{\kappa_{3O}}(\boldsymbol{s_{3O}})$). Let the original model be characterised by $M^{old} = M$, $O^{old} = O$ and U^{old}=1. Recall the structure of the full-rank reparametrised model: the cell-probabilities associated with one of the mixtures M_l are estimated and the remaining mixtures expressed as linear combinations of this mixture and the observable bases. We term the mixture M_l a *proxy basis* since it plays exactly the same role as the observable bases. The full-rank reparametrised

model can then be characterised by $M^{new} = M^{old} - 1$, $O^{new} = O^{old} + 1$ and $U^{new} = U^{old} - 1 = 0$; with M^{new} the number of remaining mixtures, O^{new} the number of observable bases (now including the original and the proxy basis) and U^{new} the number of unobservable bases within the full-rank reparametrised model. For example $\boldsymbol{\kappa_{3O}}(\boldsymbol{s_{3O}})$ is characterised by $M^{new} = 1$ (formed of M_1), $O^{new} = 4$ (formed of O_1, O_2, O_3, and M_2), and U^{new}=0.

Based on this full-rank reparametrisation, we set $M^{old} = M^{new}$, $O^{old} = O^{new}$ and $U^{old} = U^{new}$ and now set another basis as unobservable: for example, in $\boldsymbol{\kappa_{2U}}(\boldsymbol{s_{2U}})$, O_3 was set as unobservable and relabelled U_2. The model examined at this step is then characterised by $M^{new} = M^{old}$, $O^{new} = O^{old} - 1$ and U^{new}=1. This comes back to the situation examined in Section 5.3.3 with just one unobservable basis, which is easily reparametrised as full-rank. Thus, the process continues recursively until we have set the appropriate U bases to be unobservable. Note that by setting, at each step, the coefficient of the newly unobservable basis to be equal to 1 minus the other coefficients, the estimable parameter combinations at each step can also be recursively expressed, based on the full-rank reparametrisation from the previous step, using the expressions presented in Section 5.3.3. For example, in $\boldsymbol{\kappa_{2U}}(\boldsymbol{s_{2U}})$, the cell-probabilities associated to mixture M_1 are expressed as $p_i^{M1} = p_i^{M2}\beta_1 + \beta_2 p_i^{O2} + \beta_3 p_i^{O3} + (1 - \beta_1 - \beta_2 - \beta_3)p_i^{O1}$. Since $O3$ becomes unobservable, let us define $\beta_4 = 1-\beta_1-\beta_2-\beta_3$ so that β_3 can be expressed as $(1-\beta_1-\beta_2-\beta_4)$ and $p_i^{M1} = p_i^{M2}\beta_1+\beta_2 p_i^{O2}+\beta_4 p_i^{O1}+(1-\beta_1-\beta_2-\beta_4)p_i^{U2}$. This re-expression of the parameters brings us back to the exact same situation as examined in Section 5.3.3 and the general expression of estimable parameter combinations presented in that section may then be used. This result will be applied to a numerical example in Section 5.4.2.

Using this recursive approach leads to the following general result. When there are at least as many mixtures as unobservable bases ($M \geq U$), U of the M mixtures are used as proxies for the U unobservable bases whilst the

remaining $(M-U)$ mixtures are expressed as a linear combination of the observable bases and the proxy bases. This is consistent with the observation made by Yantis et al. (1991) that if only samples from the mixtures are available, and that B bases are thought to underlie these mixtures, B amongst the M mixtures should be used as surrogate bases. We also note that there are $(M+O)(K-1)$ independent cells to estimate $(O+U)(K-1)+(M-U)(O+U-1)$ parameters. It follows that M should be greater or equal to U for our general results to hold. And indeed, when looking at various examples, we found that while $M < U$, the deficiency varies with K (for example a model with $M=1$, $O=2$, $U=2$, $K=4$ presented a deficiency of 6 versus 7 when $K=5$).

5.3.5 Conclusion

From the investigations conducted, we can conclude that a model with K bins, M mixtures, B bases divided in O observable and U unobservable will present a deficiency of $U(B-1)$ for $M \geq U$ and $K \geq B$. The estimable parameters will be: the $O(K-1)$ cell probabilities of all the observable bases, the $U(K-1)$ cell-probabilities of the U mixtures used as proxy bases and finally the $(M-U)(B-1)$ coefficients relating to the $M-U$ mixtures expressed as linear combinations of the U proxies and O observable bases.

Also, intuitively, since U mixtures are being used as proxies for the unobservable bases, the mixing probabilities associated with these U mixtures will never be estimable separately. Since there are $B-1$ mixing probabilities per mixture, we come back to the result that the deficiency will be $U(B-1)$, albeit in an informal manner. Finally, we notice that if $M=U$ all the mixtures are used as proxy bases and estimated independently, a mixture-bases structure becomes apparent only when $M > U$, so this is a necessary condition to perform goodness-of-fit assessment of the hypothesised model structure.

5.4 Testing for unobservable bases in a multinomial mixture and bases framework

In this section, we derive a step-up goodness-of-fit procedure to test for unobservable bases. If independent samples of multinomial mixtures and bases are available, we first test whether the multinomial mixtures are consistent with being a mixture of the available (i.e. observable) bases using the usual test for mixtures from the MMLM approach (described in Chapter 4, Section 4.2). If this is rejected, we test for a model structure including one unobservable basis. If this is also rejected, we test for two unobservable bases.

To construct the tests for unobservable bases, we use the MMLM approach presented in Section 4.2, combined with the parameter redundancy results obtained in Sections 5.2 and 5.3 for the maximum-likelihood estimation. We will also show that it is preferable in the framework of unobservable bases to test for linear combinations rather than a mixture structure. The goodness-of-fit step is performed as usual using the statistics presented in Section 4.3 (Chapter 4): Pearson's chi-square statistic, denoted χ^2, the Cressie & Read power-divergence statistic with $\lambda = 2/3$ denoted $CR(\lambda = \frac{2}{3})$ and the log-likelihood ratio statistic G^2. The tests examined in this section correspond to goodness-of-fit assessment in terms of the reparametrised models obtained by applying the results from Section 5.3. Hence, the null hypotheses are not formulated directly in terms of the unobservable bases but rather in terms of the reparametrised model with proxy bases and these will be detailed in the relevant sections. We use simulation under very large sample size conditions in order to assess whether the procedure works as expected in theory for different scenarios: mixtures generated only by the observable bases, mixtures generated by also one unobservable basis or mixtures generated by two additional unobservable bases. As mentioned in Chapter 4, if the minimal expected value is lower than 2, the test is considered Non Applicable and denoted by NA.

5.4.1 Testing for one unobservable basis

Using the R *rmultinom* function, 300 datasets of a multinomial mixture and bases data structure were simulated, with $M = 2$, $B = 3$, $K = 7$, each multinomial being of size 6000. The true value of the multinomial cell probabilities are shown in Table 5.5. The mixtures $M1$ and $M2$ are generated by the three bases $B1$, $B2$ and $B3$ with respective mixing probabilities $\pi_1 = 0.22$, $\pi_2 = 0.36$, $\pi_3 = 0.42$ and $\gamma_1 = 0.12$, $\gamma_2 = 0.63$, $\gamma_3 = 0.25$.

To examine a situation with one unobservable basis, we set basis B3 to be unobservable (i.e. there is no sample available from B3). Proceeding as we would in a real application, we first use the test for mixtures from the original MMLM approach. The null hypothesis tested for is then "*The multinomials $M1$ and $M2$ are consistent with being mixtures of the (observed) bases $B1$ and $B2$*". The results obtained are given in Table 5.6, in terms of percentage of significant test results at a 5% level for the test-statistics considered. The test for mixtures reacts as expected: 100% of the results are significant in the presence of one unobservable basis. We then proceed to testing the fit of a model structure including one unobservable basis. Based on the results derived in Section 5.3, this equates to using the MMLM approach with one of the mixtures used as a proxy basis: its cell probabilities are estimated on their own, in the same way as the other bases and for the remaining mixtures, mixing probabilities pertaining to the proxy basis and the other bases are estimated. For example if $M2$ is used as the proxy basis, the null hypothesis tested for

Table 5.5: Multinomial cell-probabilities of mixtures and bases used to simulate datasets. The mixtures are denoted by M, the bases by B. When $B3$ is unobservable, the highest and lowest probabilities within each bin are indicated, respectively in green and red.

0.1892	**0.1682**	**0.1866**	**0.1526**	0.0966	**0.0772**	**0.1296**	M1
0.2556	0.1197	0.1906	0.1277	0.1064	0.0861	0.1139	M2
0.2300	0.1100	**0.3300**	**0.0500**	**0.0900**	**0.1300**	**0.0600**	B1
0.3500	**0.0500**	0.2000	0.0900	**0.1200**	0.1000	0.0900	B2
0.0300	0.3000	0.1000	0.2600	0.0800	0.0300	0.2000	B3

is "*Multinomial* $M1$ *is consistent with being a mixture of the (observed) bases* $B1$ *and* $B2$ *and proxy basis* $M2$". The results obtained from the test for one unobservable basis are displayed in Table 5.6. They show that the choice of the mixture used as proxy basis is actually important and can completely change the conclusions of the test. Indeed, the Type I error rate is slightly high but still close to the expected level (7.67%) when using M1 as proxy basis, however 100% of the results are significant when using M2 as proxy basis. Here, we differ from Yantis et al. (1991) who, in a more general case where the bases are not directly observable or the partition between mixture and bases unknown, advise to arbitrarily use U mixtures as surrogate bases and, then, check the fit of all possible partitions. Indeed we show, using a simple example, that choosing the mixture that will act as proxy basis cannot be done arbitrarily.

Let M1, M2 and M3 be three mixtures of the same bases B1 and B2. They are characterised as follows:

- M1: $\pi_1 B1 + \pi_2 B2$
- M2: $\nu_1 B1 + \nu_2 B2$
- M3: $\gamma_1 B1 + \gamma_2 B2$

Choosing some mixtures as proxy bases means expressing the remaining mixtures as a mixture of the observable bases and the proxy bases. The arbitrary choice of mixtures to act as proxy bases therefore boils down to the question: can all mixtures of the same bases also be expressed as mixtures of each other? Suppose that none of the bases from our example are observable. Expressing $M2$ as a mixture of $M1$ and $M3$, $M2 = \alpha_1 M1 + \alpha_2 M3$ with $\alpha_1 + \alpha_2 = 1$. It follows that

$$\begin{aligned}\nu_1 B1 + \nu_2 B2 &= \alpha_1(\pi_1 B1 + \pi_2 B2) + \alpha_2(\gamma_1 B1 + \gamma_2 B2) \\ &= \alpha_1\pi_1 B1 + \alpha_1\pi_2 B2 + \alpha_2\gamma_1 B1 + \alpha_2\gamma_2 B2\end{aligned}$$

Thus, $\nu_1 = \alpha_1\pi_1 + \alpha_2\gamma_1$, $\nu_2 = \alpha_1\pi_2 + \alpha_2\gamma_2$, and $\nu_2\gamma_1 - \nu_1\gamma_2 = \gamma_1\alpha_1\pi_2 + \gamma_1\alpha_2\gamma_2 - \alpha_1\pi_1\gamma_2 - \alpha_2\gamma_1\gamma_2$. Since all the mixing probabilities sum to 1, $\nu_1 + \nu_2 = \pi_1 + \pi_2 = \gamma_1 + \gamma_2 = 1$. It is then possible to express α_1 in terms of the original mixing probabilities:

$$\begin{aligned}\alpha_1 &= \frac{\nu_2\gamma_1 - \nu_1\gamma_2}{\gamma_1\pi_2 - \pi_1\gamma_2} \\ &= \frac{(1-\nu_1)\gamma_1 - \nu_1(1-\gamma_1)}{\gamma_1(1-\pi_1) - \pi_1(1-\gamma_1)} \\ &= \frac{\gamma_1 - \nu_1}{\gamma_1 - \pi_1}\end{aligned}$$

There will be a natural ordering between the mixing probabilities: if $0 < \pi_1 < \nu_1 < \gamma_1 < 1$ then $\gamma_1 - \nu_1 < \gamma_1 - \pi_1$ and both these quantities are positive, thus $0 < \alpha_1 < 1$, and M2 is indeed a mixture of M1 and M3.
However, if $0 < \nu_1 < \pi_1 < \gamma_1 < 1$, then $\gamma_1 - \nu_1 > \gamma_1 - \pi_1$ and $\alpha_1 > 1$. On the other hand, if $0 < \pi_1 < \gamma_1 < \nu_1 < 1$, $\gamma_1 - \nu_1 < 0$ whereas $\gamma_1 - \pi_1 > 0$, leading to $\alpha_1 < 0$. Hence, in this simple example, although M1, M2 and M3 are mixtures of the same bases B1 and B2, M2 may be expressed as a mixture of M1 and M3 only in the configuration $0 < \pi_1 < \nu_1 < \gamma_1 < 1$. The other orderings examined lead to α_1 not being between 0 and 1. Note, however, that although M2 cannot be expressed as a mixture of M1 and M3 in these configurations, it may still be expressed as a *linear combination* of those mixtures.

Although this simple example deals with a situation where none of the bases are observable, the same applies when some of the bases are observable and this can be explained intuitively for our multinomial framework. The mixture cell-probabilities are, by definition, weighted averages of the original bases cell-probabilities, thus the mixture cell-probabilities will always be bounded by the extrema of the associated bases cell-probabilities. Consequently, when one or more bases are unobservable, any mixture presenting extreme probabilities within a bin (i.e. highest or lowest probability within that bin) cannot be

modelled as a mixture of observable and proxy bases. This is illustrated in Table 5.5, where, once B3 (denoted in grey) is set as unobservable, the extreme cell-probabilities within each bin are denoted in red (minimum) and green (maximum). Mixture M1 presents extreme probability values and therefore cannot be reconstructed as a mixture of B1, B2 and M2, whereas M2 always presents intermediate cell-probabilities. Hence, only M1 may be used as a proxy basis and M2 modelled as a mixture of B1, B2 and M1. This observation is consistent with the results obtained from the mixture test for one unobservable basis in Table 5.6.

To circumvent the problem of having to choose the correct mixture as proxy basis, particularly in more complex situations, we propose to use a test for linear combinations rather than mixtures. The only difference lies in the fact that at the maximum likelihood estimation stage, coefficients are estimated rather than mixing probabilities: they are no longer constrained to lie between 0 and 1. They are, however, still constrained to sum up to 1 in accordance with the parameter redundancy results obtained in Section 5.3.

Note that we did not perform a linear combination test for 0 unobservable basis, since under the null hypothesis corresponding to this model structure, the mixtures are all "proper" mixtures of the observable bases. Thus the use of the original test for mixtures, which assesses whether the observed multinomial mixtures are consistent with being mixtures of the observable bases, is adequate. For the mixture test for one unobservable basis, once the model has been appropriately reformulated in terms of the proxy basis, recall that the null hypothesis is concerned with the structure of the remaining mixtures (not used as proxy basis): *"The remaining multinomial mixtures are consistent with being mixtures of the observed multinomial bases and the proxy basis"*. Whilst the null hypothesis for the corresponding linear combination test is: *"The remaining multinomial mixtures are consistent with being linear combinations of the observed multinomial bases and the proxy basis"*.

Table 5.6 shows that the Type I error obtained with the linear combination test (3.67%) is closer to the expected 5% level than the one obtained with the mixture test, when the correct mixture was used as proxy basis (7.67%). We also note that the linear combination tests lead to the same results, whatever mixture is chosen as proxy basis. Plotting the distribution of the Pearson's chi-square test-statistics obtained under the null hypothesis versus the theoretical chi-square distribution in Figure 5.2 shows that, for our simulations, the linear combination test presented better distributional properties in general (observed density closer to theoretical one), for the values we looked at.

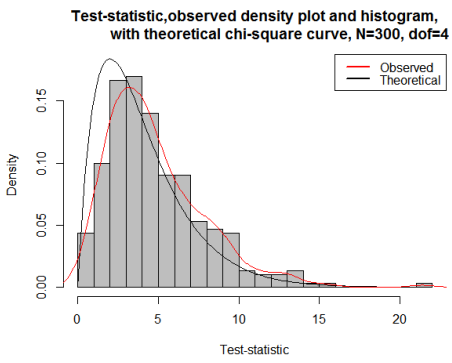
To summarise, the procedure for testing for one unobservable basis works as expected based on our simulations, and it seems preferable to test for linear combinations rather than mixtures, based on the better distributional properties of the test-statistics and the fact that any mixture chosen as proxy basis leads to the same results.

5.4.2 Testing for two unobservable bases

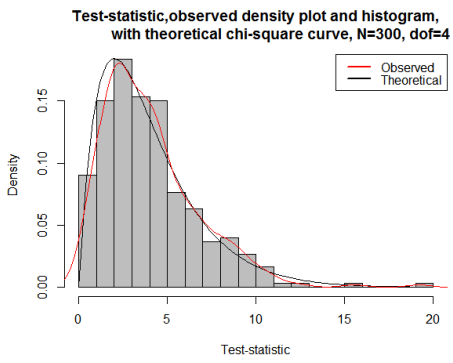
In this section, we extend the procedure to testing for two unobservable bases. Following the results from Section 5.3, at least three mixtures are needed to test the fit of a model including two unobservable bases. Two of these

Table 5.6: Procedure for testing for one unobservable basis in a general multinomial mixture and bases framework: simulation results, percentage of significant test results (5% level) out of the N applicable tests. U_{true} denotes the number of unobservable bases in the data, U_{fitted} denotes the number of unobservable states used in the fitted and tested model, M_{proxy} denotes (when applicable), the mixture used as a proxy basis. Test indicates the type of test used: M for mixture and LC for linear combination

U_{true}	U_{fitted}	M_{proxy}	χ^2(%)	CR ($\lambda = \frac{2}{3}$)(%)	G^2(%)	dof	N	Test
1	0	-	**100.00**	**100.00**	**100.00**	10	300	M
1	1	$M2$	**100.00**	**100.00**	**100.00**	4	300	M
1	1	$M1$	7.67	7.67	7.67	4	300	M
1	1	$M2$	3.67	3.67	3.67	4	300	LC
1	1	$M1$	3.67	3.67	3.67	4	300	LC



(a) Mixture test: Pearson's chi square, test-statistic distribution: $M = 2$, $O = 2$, $U = 1$, mixture $M1$ used as proxy basis



(b) Linear combination test: Pearson's chi square, test-statistic distribution: $M = 2$, $O = 2$, $U = 1$, mixture $M1$ used as proxy basis

Figure 5.2: Pearson's chi-square test-statistic distribution under the null hypothesis, tests for one unobservable basis, $M1$ used as proxy basis

mixtures are used as proxy bases and the MMLM approach is then applied to the model structure based on the remaining mixtures and the new bases (observable and proxy). For two unobservable bases, we just add a step to the procedure presented for one unobservable basis: this time we expect the original test for mixtures (0 unobservable basis) to yield a significant result, then we test for one unobservable basis, and also expect the result to be significant, finally we test for two unobservable bases and expect the test result to not be significant. Once again, we use simulation under very good conditions to verify that the procedure works as expected in theory. 300 datasets of a multinomial mixture-bases model structure with $M = 3$, $O = 2$, $U = 2$, $K = 7$, each multinomial being of size 6000, with associated cell-probabilities displayed in Table 5.7. The mixtures $M1$, $M2$, and $M3$ are generated by the four bases $B1$, $B2$, $B3$, and $B4$ with respective mixing probabilities $\pi_1 = 0.8$, $\pi_2 = 0.04$, $\pi_3 = 0.1$, $\pi_4 = 0.06$ (M1); $\gamma_1 = 0.05$, $\gamma_2 = 0.2$, $\gamma_3 = 0.6$, $\gamma_4 = 0.15$ (M2) and $\nu_1 = 0.17$, $\nu_2 = 0.15$, $\nu_3 = 0.08$, $\nu_4 = 0.6$ (M3). Bases $B3$ and $B4$ are set as unobservable. The extreme probabilities within each bin, once $B3$ and $B4$ are set as unobservable, are shown in Table 5.7 (green for the highest and red for the lowest): $M2$ and $M3$ both present extreme probabilities and

thus, should be used as proxy bases when testing for mixtures.

The results obtained using both tests for mixtures and tests for linear combinations are presented in Table 5.9. The null hypotheses associated with the tests for two unobservable bases are, once again, concerned with the structure of the remaining mixtures (i.e. those not used as proxy bases). The null hypotheses are respectively: "*The remaining multinomial mixtures are consistent with being mixtures of the observed bases and the two proxy bases*" when testing for mixtures, and "*The remaining multinomial mixtures are consistent with being linear combinations of the observed bases and the two proxy bases*" when testing for linear combinations.

As expected, the original mixture test rejects a model structure with all bases observable: the observed mixtures are not consistent with being mixtures of the observable bases; the model including one unobservable basis is also strongly rejected by the mixture tests, whatever the mixture used as a proxy basis (100% of significant test results). Finally, as expected when examining the extreme probabilities within each bin, the test for mixtures using (M1,M3) or (M1,M2) as proxy bases rejects the corresponding model structure (100% of tests rejected); whilst the Type I error rate for the mixture test for two unobservable bases using (M2, M3) as proxy bases is slightly higher than expected (8.01% of significant results).

Unexpected optimisation issues were encountered within the linear combination framework when testing for two unobservable bases. The model struc-

Table 5.7: Cell-probabilities of the multinomial mixture and bases structures used for the simulations. Once $B3$ and $B4$ are unobservable, the highest probability in each bin is denoted in green and the lowest in red.

0.720440	0.057920	0.011540	0.037380	0.087260	0.052180	0.020440	0.012840	M1
0.047350	0.004980	0.037450	0.162100	**0.480400**	0.121250	**0.116220**	0.030250	M2
0.154140	0.013954	0.030330	0.125900	0.066720	**0.481230**	0.020336	**0.107390**	M3
0.900000	**0.072000**	**0.005000**	**0.006000**	0.009000	0.005000	**0.001000**	**0.002000**	B1
0.002000	**0.003000**	**0.180000**	**0.800000**	**0.005000**	**0.002000**	0.005000	0.003000	B2
0.003000	0.000800	0.001000	0.001000	0.798000	0.001000	0.190200	0.005000	B3
0.001000	0.002000	0.004000	0.008000	0.001000	0.800000	0.007000	0.177000	B4

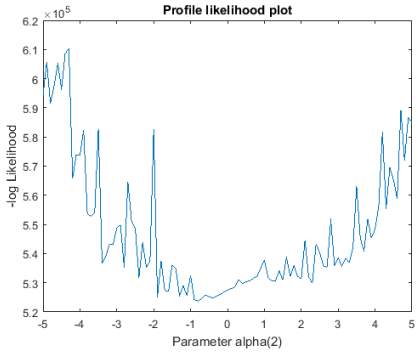
Table 5.8: Model structure estimated and tested for in the linear combination framework. The proxy bases and observable bases are denoted in red whilst the remaining mixture is denoted in blue.

$\alpha_1 p_1^{M2} + \alpha_2 p_1^{M3}$ $+\alpha_3 p_1^{B1} + (1-\alpha_1-\alpha_2-\alpha_3) p_1^{B2}$	...	$\alpha_1(1-\sum_{i=1}^{7} p_i^{M2}) + \alpha_2(1-\sum_{i=1}^{7} p_i^{M3})$ $+\alpha_3(1-\sum_{i=1}^{7} p_i^{B1}) + (1-\alpha_1-\alpha_2-\alpha_3)(1-\sum_{i=1}^{7} p_i^{B2})$	M1
p_1^{M2}	...	$(1-\sum_{i=1}^{7} p_i^{M2})$	M2
p_1^{M3}	...	$(1-\sum_{i=1}^{7} p_i^{M3})$	M3
p_1^{B1}	...	$(1-\sum_{i=1}^{7} p_i^{B1})$	B1
p_1^{B2}	...	$(1-\sum_{i=1}^{7} p_i^{B2})$	B2

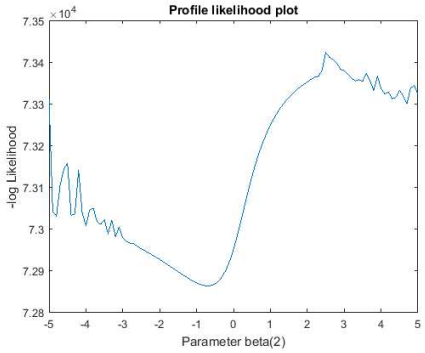
ture estimated and tested for using for example, M2 and M3 as proxy bases, is shown in Table 5.8. Different initialisations led to different optima (and thus different MLEs and test results). In order to investigate this problem, we use a dataset composed of the expected values of the multinomial cells, with each multinomial being of size 10^5. Hence the dataset is simply constructed by multiplying by 10^5 the cell probabilities from Table 5.7. For ten different initialisation starts, the following optima were reached, given in terms of -(log-likelihood) values: {523063.05, 532925.53, 546208.16, 549730.75, **522709.32**, 537286.65, 528127.90, 533394.63, 559820.20, 525353.00}. As illustrated in this example, the optimisation never reaches the same point and overall, the minimum reached is 522709.32. However, the value of -(log-likelihood) at the true parameter values is **522643.27** and this is never reached. Using even an extreme number of random initial starts as high as 10^4, the optimum reached is 522643.43 and it is reached only once out of 10^4: clearly, the optimiser keeps getting stuck in local minima. Note that although we had used the same process to optimise the likelihood in the case of one unobservable basis: $O = 2$, $U = 1$, $M = 2$, we did not encounter this issue previously. To explore this further, the "apparent" profile likelihoods of one of the coefficients are plotted in Figure 5.3 for each of the situations: $U = 1$ and $U = 2$. This is done by incrementally fixing the parameter of interest and performing maximum-likelihood estimation with respect to all the other parameters. The optimum value of the -(log-likelihood) reached is then plotted against the parameter values. Note

the use of the term "apparent" because despite a high number of initial starts, the profile likelihood also gets stuck in local minima. We observe that the likelihood surface seems very bumpy and difficult to optimise for $U = 2$ compared to $U = 1$. We perform further examinations to determine whether the peaks and troughs are indeed due to local minima, or rather, whether they reflect convergence issues. In order to do this, we plot a profile likelihood by incrementally increasing the value of the parameter of interest, and another one by incrementally decreasing this value. Also, we use random starting points for the initial optimisation, but for all the subsequent ones, we use the maximum likelihood estimates obtained in the previous step as an informative starting point. The profile likelihoods obtained are plotted in Figure 5.4. We observe that the plots are much smoother and that the troughs do not occur at the same place in both graphs. Hence, there is clearly a problem of convergence.

Due to the problems encountered, we explored an alternative approach to the optimisation process. Rather than directly optimising the log-likelihood, we used an iterative approach for the estimation based on the original algorithm proposed by Yantis et al. (1991). First the bases cell-probabilities are estimated, then the mixing probabilities updated based on these estimations, the updated mixing probabilities in turn feed in to obtain new estimates for the bases probabilities and so forth until convergence. We use this iterative process, with the difference that we estimate coefficients rather than mixing probabilities. We implemented this optimisation in Matlab and the associated pseudo-algorithm is given below. Using the iterative method on the same example dataset composed of expected multinomial cell values corresponding to the probabilities of Table 5.7, with 10 random initial starts, the optima reached are {522643.27, 522643.27, 579237.58, 522643.88, 522643.27, 522643.27, 522643.27, 522643.27, 522643.27}. The true minimum is now reached 8 times out of 10, which is very promising. We therefore choose to use the iterative method to estimate the parameters needed for our linear

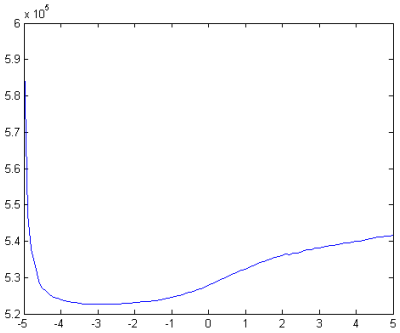


(a) α_2 profile likelihood, $M = 3$, $O = 2$, $U = 2$, mixtures $M1$ and $M2$ used as proxy bases

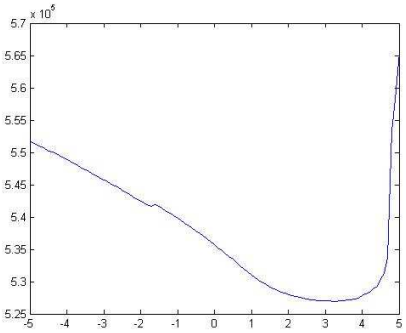


(b) β_2, $M = 2$, $O = 2$, $U = 1$ profile likelihood, $M2$ used as proxy basis

Figure 5.3: Profile likelihood for coefficients α_2 (from a model with 2 unobservable bases) and β_2 (from a model with one unobservable basis)



(a) α_2 profile likelihood obtained by incrementally increasing the parameter value, $M = 3$, $O = 2$, $U = 2$, mixtures $M1$ and $M2$ used as proxy bases



(b) α_2 profile likelihood obtained by incrementally decreasing the parameter value, $M = 3$, $O = 2$, $U = 2$, mixtures $M1$ and $M2$ used as proxy bases

Figure 5.4: Profile likelihood for coefficient α_2, $M = 3$, $O = 2$, $U = 2$, mixtures $M1$ and $M2$ used as proxy bases

Algorithm 1 Iterative optimisation for parameter estimation in the linear combination framework

procedure ITERATIVE(B, M_p, M)
  Step 1 Initialisation of basis and proxy basis cell-probabilities (here using classical multinomial MLE result)
  for $i \leftarrow 1, n_B$ **do**
    for $j \leftarrow 1, (K-1)$ **do**
      $p_B^{old}(i,j) = \frac{B(i,j)}{\sum_{k=1}^{K} B(i,k)}$
    end for
    $p_B^{old}(i,K) = 1 - \sum_{k=1}^{K-1} B(i,k)$
  end for
  for $i \leftarrow 1, n_{M_p}$ **do**
    for $j \leftarrow 1, (K-1)$ **do**
      $p_{M_p}^{old}(i,j) = \frac{M_p(i,j)}{\sum_{k=1}^{K} M_p(i,k)}$
    end for
    $p_{M_p}^{old}(i,K) = 1 - \sum_{k=1}^{K-1} M_p(i,k)$
  end for

  Step 2 Initialisation of coefficients
  for $l \leftarrow 1, 3$ **do** ▷ (3 coefficients for $M = 3, O = 2, U = 2$)
    $\alpha^{old}(l)$ =random values from normal distribution (for example)
  end for

  Step 3 Optimize (used Matlab *fminunc*) with respect to the coefficients (basis probabilities fixed to p_B^{old}, $p_{M_p}^{old}$).
Output: α^{new}, $devM_{new}$ ▷ $devM_{new}$ denotes the deviance optimum reached at this step
  Step 4 Optimize (used Matlab *fminunc*) with respect to basis and proxy basis probabilities (coefficients fixed to α^{new}) .
Output: p_B^{new},$p_{M_p}^{new}$, $devB_{new}$
  Step 5 Set $\alpha^{old} = \alpha^{new}$, $p_B^{old} = p_B^{new}$, $p_{M_p}^{old} = p_{M_p}^{new}$

  while $|devM_{new} - devB_{new}| \geq 10^{-4}$ **do**
    do **Step 3**, using α^{old} as initial values, with basis probabilities fixed to p_B^{new}, and proxy basis probabilities to $p_{M_p}^{new}$
Output: α^{new}, $devM_{new}$
    if $devM_{new} > devB_{new}$ **then**
      $\alpha^{new} = \alpha^{old}$ ▷ The new coefficient estimates do not improve the likelihood, therefore the old estimates are retained
      break
    end if
    do **Step 4**, using p_B^{old}, $p_{M_p}^{old}$ as initial values, coefficients fixed to α^{new}.
Output: p_B^{new}, $p_{M_p}^{new}$
    if $devB_{new} > devM_{new}$ **then**
      $p_B^{new} = p_B^{old}$, $p_M^{new} = p_M^{old}$ ▷ The new basis and proxy basis probability estimates do not improve the likelihood, therefore the old estimates are retained
      break
    end if
    Start a new cycle: $p_B^{old} = p_B^{new}$, $p_{M_p}^{old} = p_{M_p}^{new}$, $\alpha^{old} = \alpha^{new}$
  end while
end procedure

combination tests; and the results obtained from the subsequent goodness-of-fit assessment for the simulated datasets with two unobservable bases are given in Table 5.9. The goodness-of-fit tests for linear combinations strongly reject a model structure including one unobservable basis: the observed mixtures are not consistent with being linear combinations of the proxy basis and the observed bases (100% of significant results, whatever proxy bases is chosen). The linear combination tests for two unobservable bases yield a Type I error close to 5%, whatever proxy bases are chosen. (As a precaution, the optimisation and goodness-of-fit were also re-run using the iterative method for the simulated datasets from Section 5.4.1, when testing for linear combinations when one basis is unobservable and the results obtained were unchanged.)

Although the test results are as expected, we note that different choices of proxy basis lead to slightly different results with the linear combination approach, whilst they should all yield exactly the same result (since the cell-probabilities computed or estimated should be equal in all three cases).

Hence, we examine the maximum-likelihood estimates obtained for the coefficients, in each configuration of proxy bases and compare them to the true parameter values. In order to compute these true values, we need to express the coefficients as a function of the original mixing probabilities. We use two different approaches to do so: reformulating each mixture as a linear combination and solving the relevant equations using the Matlab symbolic algebra package, or alternatively using the recursive reparametrisation results stated in Section 5.3.4.

Recall that M1, M2 and M3 are all mixtures of B1,B2, B3 and B4 that may be expressed as:

- $M1 = \pi_1 B1 + \pi_2 B2 + \pi_3 B3 + (1 - \pi_1 - \pi_2 - \pi_3) B4$,
- $M2 = \gamma_1 B1 + \gamma_2 B2 + \gamma_3 B3 + (1 - \gamma_1 - \gamma_2 - \gamma_3) B4$
- $M3 = \nu_1 B1 + \nu_2 B2 + \nu_3 B3 + (1 - \nu_1 - \nu_2 - \nu_3) B4$,

Table 5.9: Procedure testing for two unobservable bases in a general multinomial mixture and bases framework, simulation results, percentage of significant test results (5% level) out of N, the number of applicable tests. U_{true} denotes the number of unobservable bases in the simulated data, U_{fitted} denotes the number of unobservable bases used in the fitted and tested model, M_{proxy} denotes (when applicable), the mixture(s) used as proxy bases. Test M indicates the use of a mixture test and $LC(it)$ the use of a linear combination test based on the iterative approach for parameter estimation.

U_{true}	U_{fitted}	M_{proxy}	$\chi^2(\%)$	CR ($\lambda = \frac{2}{3}$)(%)	$G^2(\%)$	dof	N	Test
2	0	-	**100.00**	**100.00**	**100.00**	18	235	M
2	1	$M3$	**100.00**	**100.00**	**100.00**	10	287	M
2	1	$M2$	**100.00**	**100.00**	**100.00**	10	291	M
2	1	$M1$	**100.00**	**100.00**	**100.00**	10	213	M
2	2	$M2$, $M3$	8.01	8.01	8.01	4	287	M
2	2	$M1$, $M3$	**100.00**	**100.00**	**100.00**	4	282	M
2	2	$M1$, $M2$	**100.00**	**100.00**	**100.00**	4	245	M
2	1	$M3$	**100.00**	**100.00**	**100.00**	10	294	LC(it)
2	1	$M2$	**100.00**	**100.00**	**100.00**	10	294	LC(it)
2	1	$M1$	**100.00**	**100.00**	**100.00**	10	293	LC(it)
2	2	$M2$, $M3$	3.83	3.83	3.48	4	287	LC(it)
2	2	$M1$, $M2$	5.28	5.28	5.28	4	284	LC(it)
2	2	$M1$, $M3$	3.89	3.89	3.89	4	283	LC(it)

Expressing $M1$ as a linear combination of the observable bases $B1$, $B2$ and the mixtures $M2$,$M3$, with the coefficients summing to one, results in:

$$
\begin{aligned}
M1 &= \alpha_1 M3 + \alpha_2 M2 + \alpha_3 B1 + (1 - \alpha_1 - \alpha_2 - \alpha_3) B2 \\
&= \alpha_1 \left[\nu_1 B1 + \nu_2 B2 + \nu_3 B3 + (1 - \nu_1 - \nu_2 - \nu_3) B4\right] \\
&+ \alpha_2 \left[\gamma_1 B1 + \gamma_2 B2 + \gamma_3 B3 + (1 - \gamma_1 - \gamma_2 - \gamma_3) B4\right] \\
&+ \alpha_3 B1 + (1 - \alpha_1 - \alpha_2 - \alpha_3)\, B2 \\
&= (\alpha_1 \nu_1 + \alpha_2 \gamma_1 + \alpha_3) B1 + (\alpha_1 \nu_2 + \alpha_2 \gamma_2 + 1 - \alpha_1 - \alpha_2 - \alpha_3) B2 \\
&+ (\alpha_1 \nu_3 + \alpha_2 \gamma_3) B3 + \left[\alpha_1 (1 - \nu_1 - \nu_2 - \nu_3) + \alpha_2 (1 - \gamma_1 - \gamma_2 - \gamma_3)\right] B4
\end{aligned}
$$

Using the symbolic package from Matlab, we solve the following equations for $\alpha_1, \alpha_2, \alpha_3$

$$\begin{cases} \pi_1 = \alpha_1\nu_1 + \alpha_2\gamma_1 + \alpha_3 \\ \pi_2 = \alpha_1\nu_2 + \alpha_2\gamma_2 + 1 - \alpha_1 - \alpha_2 - \alpha_3 \\ \pi_3 = \alpha_1\nu_3 + \alpha_2\gamma_3 \\ 1 - \pi_1 - \pi_2 - \pi_3 = \alpha_1(1 - \nu_1 - \nu_2 - \nu_3) + \alpha_2(1 - \gamma_1 - \gamma_2 - \gamma_3) \end{cases}$$

This results in:

$$\alpha_1 = -\frac{\pi_3 - \gamma_3 + \pi_1\gamma_3 - \pi_3\gamma_1 + \pi_2\gamma_3 - \pi_3\gamma_2}{\gamma_3 - \nu_3 + \gamma_1\nu_3 - \gamma_3\nu_1 + \gamma_2\nu_3 - \gamma_3\nu_2}$$
$$\alpha_2 = \frac{\pi_3 - \nu_3 + \pi_1\nu_3 - \pi_3\nu_1 + \pi_2\nu_3 - \pi_3\nu_2}{\gamma_3 - \nu_3 + \gamma_1\nu_3 - \gamma_3\nu_1 + \gamma_2\nu_3 - \gamma_3\nu_2}$$
$$\alpha_3 = \frac{\pi_1\gamma_3 - \pi_3\gamma_1 - \pi_1\nu_3 + \pi_3\nu_1 + \gamma_1\nu_3 - \gamma_3\nu 1 + \pi_1\gamma_2\nu_3 - \pi_1\gamma_3\nu_2 - \pi_2\gamma_1\nu_3 + \pi_2\gamma_3\nu_1 + \pi_3\gamma_1\nu_2 - \pi_3\gamma_2\nu_1}{\gamma_3 - \nu_3 + \gamma_1\nu_3 - \gamma_3\nu_1 + \gamma_2\nu_3 - \gamma_3\nu_2}$$

Results for the other configurations of proxy bases can be obtained in the same way: M3 as a function of M1 and M2 results in

$M3 = \beta_1 M1 + \beta_2 M2 + \beta_3 B1 + (1 - \beta_1 - \beta_2 - \beta_3)B2$, with

$$\beta_1 = -\frac{\nu_3 - \gamma_3 + \nu_1\gamma_3 - \nu_3\gamma_1 + \nu_2\gamma_3 - \nu_3\gamma_2}{\gamma_3 - \pi_3 + \gamma_1\pi_3 - \gamma_3\pi_1 + \gamma_2\pi_3 - \gamma_3\pi_2}$$
$$\beta_2 = \frac{\nu_3 - \pi_3 + \nu_1\pi_3 - \nu_3\pi_1 + \nu_2\pi_3 - \nu_3\pi_2}{\gamma_3 - \pi_3 + \gamma_1\pi_3 - \gamma_3\pi_1 + \gamma_2\pi_3 - \gamma_3\pi_2}$$
$$\beta_3 = \frac{\nu_1\gamma_3 - \nu_3\gamma_1 - \nu_1\pi_3 + \nu_3\pi_1 + \gamma_1\pi_3 - \gamma_3\pi 1 + \nu_1\gamma_2\pi_3 - \nu_1\gamma_3\pi_2 - \nu_2\gamma_1\pi_3 + \nu_2\gamma_3\pi_1 + \nu_3\gamma_1\pi_2 - \nu_3\gamma_2\pi_1}{\gamma_3 - \pi_3 + \gamma_1\pi_3 - \gamma_3\pi_1 + \gamma_2\pi_3 - \gamma_3\pi_2}$$

M2 as a function of M1 and M3 results in :

$M2 = \tau_1 M3 + \tau_2 M1 + \tau_3 B1 + (1 - \tau_1 - \tau_2 - \tau_3)B2$, with

$$\tau_1 = -\frac{\gamma_3 - \pi_3 + \gamma_1\pi_3 - \gamma_3\pi_1 + \gamma_2\pi_3 - \gamma_3\pi_2}{\pi_3 - \nu_3 + \pi_1\nu_3 - \pi_3\nu_1 + \pi_2\nu_3 - \pi_3\nu_2}$$
$$\tau_2 = \frac{\gamma_3 - \nu_3 + \gamma_1\nu_3 - \gamma_3\nu_1 + \gamma_2\nu_3 - \gamma_3\nu_2}{\pi_3 - \nu_3 + \pi_1\nu_3 - \pi_3\nu_1 + \pi_2\nu_3 - \pi_3\nu_2}$$
$$\tau_3 = \frac{\gamma_1\pi_3 - \gamma_3\pi_1 - \gamma_1\nu_3 + \gamma_3\nu_1 + \pi_1\nu_3 - \pi_3\nu 1 + \gamma_1\pi_2\nu_3 - \gamma_1\pi_3\nu_2 - \gamma_2\pi_1\nu_3 + \gamma_2\pi_3\nu_1 + \gamma_3\pi_1\nu_2 - \gamma_3\pi_2\nu_1}{\pi_3 - \nu_3 + \pi_1\nu_3 - \pi_3\nu_1 + \pi_2\nu_3 - \pi_3\nu_2}$$

The coefficients' numeric values obtained for our simulation settings, in these different configurations, are displayed in Table 5.10.

We can also use the recursive reparametrisation approach mentioned in Section 5.3.4 to obtain the coefficient values. Indeed, in a first step, B4 is rendered unobservable and M2 used as a proxy basis. M1 and M3 are then expressed as: M1=α_1^{M1}M2+α_2^{M1}B2+α_3^{M1}B3+α_4^{M1}B1 and

M3=α_1^{M2}M2+α_2^{M2}B2+α_3^{M2}B3+α_4^{M2}B1, with $\alpha_4^{Mi} = 1 - \alpha_1^{Mi} - \alpha_2^{Mi} - \alpha_3^{Mi}$.

From the general expression of the estimable combinations obtained in Section 5.3.3, $\alpha_1^{M1} = \frac{1-\pi_1-\pi_2-\pi_3}{1-\gamma_1-\gamma_2-\gamma_3}$, $\alpha_2^{M1} = \frac{\pi_2(-1+\gamma_1+\gamma_3)+\gamma_2(1-\pi_1-\pi_3)}{-(1-\gamma_1-\gamma_2-\gamma_3)}$,
$\alpha_3^{M1} = \frac{\pi_3(-1+\gamma_1+\gamma_2)+\gamma_3(1-\pi_1-\pi_2)}{-(1-\gamma_1-\gamma_2-\gamma_3)}$, $\alpha_4^{M1} = 1 - \alpha_1^{M1} - \alpha_2^{M1} - \alpha_3^{M1}$
and $\alpha_1^{M3} = \frac{1-\nu_1-\nu_2-\nu_3}{1-\gamma_1-\gamma_2-\gamma_3}$, $\alpha_2^{M3} = \frac{\nu_2(-1+\gamma_1+\gamma_3)+\gamma_2(1-\nu_1-\nu_3)}{-(1-\gamma_1-\gamma_2-\gamma_3)}$,
$\alpha_3^{M3} = \frac{\nu_3(-1+\gamma_1+\gamma_2)+\gamma_3(1-\nu_1-\nu_2)}{-(1-\gamma_1-\gamma_2-\gamma_3)}$, $\alpha_4^{M3} = 1 - \alpha_1^{M3} - \alpha_2^{M3} - \alpha_3^{M3}$

This reparametrised model is full rank. We then render B3 unobservable and express its coefficient as one minus the others: $\alpha_3^{Mi} = 1 - \alpha_1^{Mi} - \alpha_2^{Mi} - \alpha_4^{Mi}$ and we repeat the process, using M1 as additional proxy basis. Hence, M3 can be expressed as M3=β_1^{M3}M1+β_2^{M3}M2+β_3^{M3}B2+β_4^{M3}B1, with
$\beta_1^{M3} = \frac{1-\alpha_1^{M3}-\alpha_2^{M3}-\alpha_4^{M3}}{1-\alpha_1^{M1}-\alpha_2^{M1}-\alpha_4^{M1}}$, $\beta_2^{M3} = \frac{\alpha_1^{M2}(-1+\alpha_2^{M1}+\alpha_4^{M1})+\alpha_1^{M1}(1-\alpha_2^{M2}-\alpha_4^{M2})}{-(1-\alpha_1^{M1}-\alpha_2^{M1}-\alpha_4^{M1})}$,
$\beta_3^{M3} = \frac{\alpha_2^{M2}(-1+\alpha_1^{M1}+\alpha_4^{M1})+\alpha_2^{M1}(1-\alpha_1^{M2}-\alpha_4^{M2})}{-(1-\alpha_1^{M1}-\alpha_2^{M1}-\alpha_4^{M1})}$, and $\beta_4^{M3} = 1-\beta_1^{M3}-\beta_2^{M3}-\beta_3^{M3}$.Finally, in the last step, the α terms can be replaced by their expression as a function of the original mixing probabilities (note that the Matlab symbolic package was used to simplify the expressions):

$\beta_1^{M3} = \frac{\gamma_3(\nu_1+\nu_2-1)-\nu_3(\gamma_1+\gamma_2-1)}{\gamma_3(\pi_1+\pi_2-1)-\pi_3(\gamma_1+\gamma_2-1)}$, $\beta_2^{M3} = -\frac{\nu_3-\pi_3+\nu_1\pi_3-\nu_3\pi_1+\nu_2\pi_3-\nu_3\pi_2}{-(\gamma_3-\pi_3+\gamma_1\pi_3-\gamma_3\pi_1+\gamma_2\pi_3-\gamma_3\pi_2)}$
$\beta_3^{M3} = \frac{\gamma_2\nu_3-\gamma_3\nu_2-\gamma_2\pi_3+\gamma_3\pi_2+\nu_2\pi_3-\nu_3\pi_2-\gamma_1\nu_2\pi_3+\gamma_1\nu_3\pi_2+\gamma_2\nu_1\pi_3-\gamma_2\nu_3\pi_1-\gamma_3\nu_1\pi_2+\gamma_3\nu_2\pi_1}{-(\gamma_3-\pi_3+\gamma_1\pi_3-\gamma_3\pi_1+\gamma_2\pi_3\gamma_3\pi_2)}$,
$\beta_4^{M3} = \frac{\gamma_1\nu_3-\gamma_3\nu_1-\gamma_1\pi_3+\gamma_3\pi_1+\nu_1\pi_3-\nu_3\pi_1+\gamma_1\nu_2\pi_3-\gamma_1\nu_3\pi_2-\gamma_2\nu_1\pi_3+\gamma_2\nu_3\pi_1+\gamma_3\nu_1\pi_2-\gamma_3\nu_2\pi_1}{-(\gamma_3-\pi_3+\gamma_1\pi_3-\gamma_3\pi_1+\gamma_2\pi3-\gamma_3\pi_2)}$.

Comparing to the results previously obtained using a direct reformulation as linear combination and solving the corresponding equations, $\beta_1^{M3} = \beta_1$, $\beta_2^{M3} = \beta_2$ and $\beta_4^{M3} = \beta_3$; both methods give the same results.

The values of the coefficients presented in Table 5.10 indicate that even when the mixtures M2 and M3 (which present extreme probabilities within the bins) are chosen as proxy bases, M1 cannot be expressed as a mixture of these and the observable bases: indeed the coefficient relating to B2 is negative (-0.0007). This shows that there can be situations in which none of the mixtures can be expressed strictly as a mixture of proxy and observable bases; and this is another argument in favour of using linear combination tests rather than mixture tests.

Table 5.10: Coefficient values for the different configurations of proxy bases

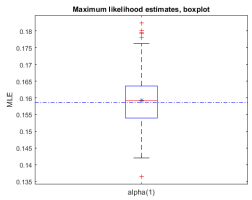
Mixture	Bases and proxy bases				
	$M1$	$M3$	$M2$	$B1$	$B2$
M1	-	0.1586	0.0603	0.7818	-0.0007
M2	16.5714	-2.6286	-	-12.9557	0.0129
M3	6.3043	-	-0.3804	-4.9288	0.0049

Boxplots of the maximum likelihood estimates of the coefficients, obtained using the iterative optimisation method, are shown in Figures 5.5 to 5.7 for the different configuration of proxy bases. We observe from Figures 5.6 and 5.7 that some of the estimates are biased. This might account for the slight difference in the test results. Although the tests only seem slightly affected by this bias, we try using a stricter convergence criterion and fix it to 10^{-9} instead of 10^{-4}. This more stringent criterion increased computational time (from hours to days). Due to this, results were generated for only one of the configurations: M1 and M3 used as proxy bases; and the boxplot of the MLEs of interest are shown in Figure 5.8. These boxplots show that the stricter convergence criterion produces very good results and should be retained.

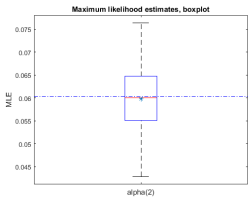
To briefly summarise, the procedure for testing for two unobservable bases works as expected in theory based on our simulations. The tests used for unobservable bases should be linear combination tests and the estimation step should be performed using the iterative optimisation method with a stringent enough convergence criterion (e.g. 10^{-9}). When using these tests, the choice of original mixtures used as proxy bases does not matter.

5.5 Application: testing for unobservable states in a capture-recapture setting

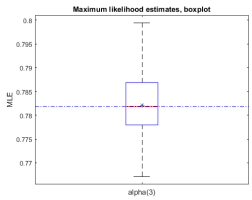
Returning to our initial capture-recapture setting, recall from Section 5.1 that an unobservable state translates itself into an unobservable basis within a multinomial mixture and bases framework. Once animals are grouped into



(a) Maximum likelihood estimates boxplot, coefficient of $M2$

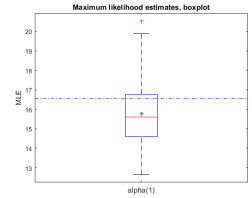


(b) Maximum likelihood estimates boxplot, coefficient of $M3$

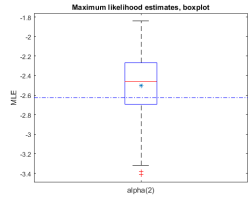


(c) Maximum likelihood estimates boxplot, coefficient of $B1$

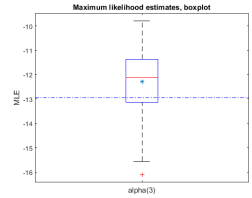
Figure 5.5: Coefficients maximum likelihood estimates, boxplots, iterative optimization: $M2, M3$ proxy basis. The blue horizontal dashed line represents the true value of the coefficient and $*$ represents the mean.



(a) Maximum likelihood estimates boxplot, coefficient of $M1$

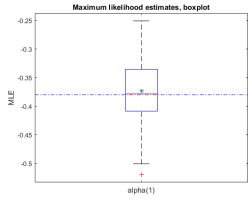


(b) Maximum likelihood estimates boxplot, coefficient of $M3$

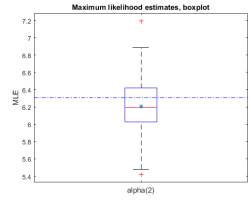


(c) Maximum likelihood estimates boxplot, coefficient of $B1$

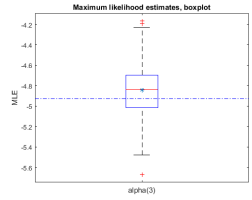
Figure 5.6: Coefficients maximum likelihood estimates boxplots, iterative optimization: $M1, M3$ proxy basis. The blue horizontal dashed line represents the true value of the coefficient and $*$ represents the mean.



(a) Maximum likelihood estimates boxplot, coefficient of $M2$

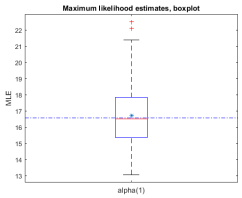


(b) Maximum likelihood estimates boxplot, coefficient of $M1$

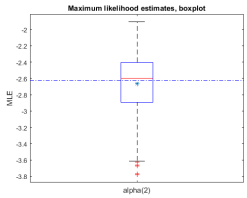


(c) Maximum likelihood estimates boxplot, coefficient of $B1$

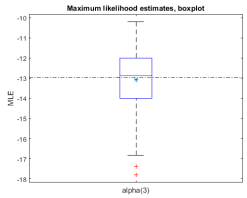
Figure 5.7: Coefficients maximum likelihood estimates boxplots, iterative optimization: $M1, M2$ proxy basis. The blue horizontal dashed line represents the true value of the coefficient and $*$ represents the mean.



(a) Maximum likelihood estimates boxplot, coefficient of $M1$



(b) Maximum likelihood estimates boxplot, coefficient of $M3$



(c) Maximum likelihood estimates boxplot, coefficient of $B1$

Figure 5.8: Coefficients, boxplots of maximum likelihood estimates obtained with the iterative optimisation, using a more stringent convergence criterion (10^{-9}), $M1, M3$ proxy basis. The blue horizontal dashed line represents the true value of the coefficient and $*$ represents the mean..

the relevant m-array cells used to perform the existing Test M, the resulting table consists of independent samples from multinomial mixtures and their associated bases, with unobservable states corresponding to unobservable bases. In other words, once the table required for Test M is constructed, we enter the exact same framework of mixtures and bases that has been analysed throughout this chapter. Hence, based on the results obtained in Sections 5.2 to 5.4, we use the procedure based on linear combination tests, with the parameter estimation step performed using the iterative optimisation method (convergence criterion set to 10^{-9}), to examine whether this procedure can work in a capture-recapture framework, for testing for unobservable states. Due to the added complexity resulting from the fact that mixture and bases cell-probabilities are formed by products of capture-recapture parameters, we use simulation under very large sample size conditions to examine whether the procedure can work in a capture-recapture setting.

5.5.1 Testing for one unobservable state

300 datasets are simulated from a capture-recapture experiment with 5 sampling occasions, with 25,000 animals released per occasion. The animals move between three live states but only two are observable, all the parameters are constant with time. Using the general multievent notations introduced in

Chapter 4, the model is defined by $\boldsymbol{\Pi_t} = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \end{bmatrix}$;

$\boldsymbol{\Phi_t} = \begin{bmatrix} 0.7 & 0 & 0 & 0.3 \\ 0 & 0.8 & 0 & 0.2 \\ 0 & 0 & 0.9 & 0.1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$; $\boldsymbol{\Psi_t} = \begin{bmatrix} 0.1 & 0.3 & 0.6 & 0 \\ 0.3 & 0.15 & 0.55 & 0 \\ 0.4 & 0.4 & 0.2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ for $t = 1, \ldots, 4$ and

$\boldsymbol{B_t} = \begin{bmatrix} 0.45 & 0.55 & 0 \\ 0.45 & 0 & 0.55 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ for $t = 2, \ldots, 5$

The results obtained for the procedure based on tests for linear combinations are given in Table 5.11. Recall that Test M is used to test for a model structure with no unobservable states. Test M is significant at occasions 2 and 3 (100% of significant results): the animals not captured at time i are not consistent with being a mixture of the animals captured at time i in states 1 and 2. The linear combination test for one unobservable state presents a Type I error rate close to the expected 5% at both occasions 2 and 3 and the results are the same, whatever proxy mixture is chosen. Hence, these simulations show that the procedure for testing for one unobservable basis can be used in a capture-recapture setting, to test for one unobservable state.

5.5.2 Testing for two unobservable states

We now attempt to test for two unobservable states in a capture-recapture framework by applying the procedure retained to test for two unobservable bases. Since our main aim is to assess whether the procedure can work in a capture-recapture context, we focus on simulations generated by an extreme situation with a model structure which ensures that the mixtures at occasions 2 and 3 cannot be reconstructed solely from the observable bases nor from including just one unobservable basis in the model structure: 300 datasets

Table 5.11: Test for detecting 1 unobservable state in a capture-recapture framework, simulation results, percentage of significant test results (5% level) out of N applicable tests. U_{true} denotes the number of unobservable states in the simulated data, U_{fitted} denotes the number of unobservable states used in the fitted and tested model, M_{proxy} denotes (when applicable), the mixture used as a proxy basis, i denotes the capture occasion.

U_{true}	U_{fitted}	M_{proxy}	χ^2(%)	CR ($\lambda = \frac{2}{3}$)(%)	G^2(%)	i	dof	N	Test
1	0	NA	**100.00**	**100.00**	**100.00**	2	8	300	M
1	0	NA	**100.00**	**100.00**	**100.00**	3	4	300	M
1	1	$M2$	4.00	4.00	4.00	2	3	300	LC(it)
1	1	$M2$	6.00	6.00	6.00	3	1	300	LC(it)
1	1	$M1$	4.00	4.00	4.00	2	3	300	LC(it)
1	1	$M1$	6.00	6.00	6.00	3	1	300	LC(it)

were simulated for a capture-recapture experiment with 7 sampling occasions with 25,000 animals are released at each occasion, the animals move between four live states,amongst which only two are observable. The probabilities were chosen in order to give rise to two very different unobservable bases, without which the mixture probabilities could not be reconstructed (at least at occasions 2 and 3). Recall that the procedure for testing for two unobservable bases requires at least three mixtures and that in a capture-recapture framework, the table used to perform Test M (see Section 4.2 from Chapter 4) contains as many mixtures as observable states, since it separates animals not captured at occasion i only by their state at release. Hence if there are two observable states, there will be two corresponding mixtures formed by the live animals not seen at i but previously released in state A and those not seen at i but previously released in state B. In order to have three mixtures, we first keep the mixtures separated by state and time of previous release and then pool some of them (arbitrarily) to obtain three mixtures. An example is given in Table 5.12: we chose to pool M1 and M3 for a test at occasion 3, whilst the table used for Test M is given in Table 5.13. The manner in which the mixtures are pooled should not affect the results due to the property from Pradel et al. (2003) cited in Section 4.3.1 (the distribution of a sum of mixtures from the

same bases is still a mixture of the same bases) and due to the fact that we are dealing with large sample sizes. For smaller datasets, the pooling decision might be a function of the data sparseness. Note that, due to the need for 3 mixtures, the linear combination test for two unobservable states may only be performed on occasions 3 to T-3. We chose to present the test results at each occasion only for the occasions common to all steps of the testing procedure (i.e. 3 to T-3), but recall that both Test M and the test for one unobservable state can be performed from occasions 2 to T-2.

The model used for our simulations is defined by $\mathbf{\Pi_t} = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 \end{bmatrix}$;

$$\mathbf{\Phi_t} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ for } t = 1, \ldots, 3$$

$$\mathbf{\Phi_t} = \begin{bmatrix} 0.25 & 0 & 0 & 0 & 0.75 \\ 0 & 0.33 & 0 & 0 & 0.67 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ for } t = 4, \ldots, 6.$$

Table 5.12: An example table used to test for two unobservable states: these mixtures are pooled arbitrarily to form the three mixtures necessary for the test.

m_{14}^{AA}	m_{14}^{AB}	m_{15}^{AA}	m_{15}^{AB}	M1
m_{14}^{BA}	m_{14}^{BB}	m_{15}^{BA}	m_{15}^{BB}	M2
m_{24}^{AA}	m_{24}^{AB}	m_{25}^{AA}	m_{25}^{AB}	M3
m_{24}^{BA}	m_{24}^{BB}	m_{25}^{BA}	m_{25}^{BB}	M4
m_{34}^{AA}	m_{34}^{AB}	m_{35}^{AA}	m_{35}^{AB}	B2
m_{34}^{BA}	m_{34}^{BB}	m_{35}^{BA}	m_{35}^{BB}	B1

Table 5.13: Table associated to the original Test M at occasion $i = 3$, an example, the mixtures are denoted by M and the bases by B.

$m_{14}^{AA} + m_{24}^{AA}$	$m_{14}^{AB} + m_{24}^{AB}$	$m_{15}^{AA} + m_{25}^{AA}$	$m_{15}^{AB} + m_{25}^{AB}$	M1
$m_{14}^{BA} + m_{24}^{BA}$	$m_{14}^{BB} + m_{24}^{BB}$	$m_{15}^{BA} + m_{25}^{BA}$	$m_{15}^{BB} + m_{25}^{BB}$	M2
m_{34}^{AA}	m_{34}^{AB}	m_{35}^{AA}	m_{35}^{AB}	B2
m_{34}^{BA}	m_{34}^{BB}	m_{35}^{BA}	m_{35}^{BB}	B1

The transition matrix is $\boldsymbol{\Psi_t} = \begin{bmatrix} 0.1 & 0.6 & 0.22 & 0.08 & 0 \\ 0.6 & 0.1 & 0.08 & 0.22 & 0 \\ 0.05 & 0.05 & 0.8 & 0.1 & 0 \\ 0.05 & 0.05 & 0.1 & 0.8 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ for $t = 1$,

$$\boldsymbol{\Psi_t} = \begin{bmatrix} 0.8 & 0.2 & 0 & 0 & 0 \\ 0.2 & 0.8 & 0 & 0 & 0 \\ 0.05 & 0.05 & 0.8 & 0.1 & 0 \\ 0.05 & 0.05 & 0.1 & 0.8 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ for } t = 2, 3$$

$$\boldsymbol{\Psi_t} = \begin{bmatrix} 0.8 & 0.2 & 0 & 0 & 0 \\ 0.2 & 0.8 & 0 & 0 & 0 \\ 0.85 & 0.15 & 0 & 0 & 0 \\ 0 & 0.75 & 0 & 0.25 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ for } t = 4 \ldots 6$$

and the event matrix $\boldsymbol{B_t} = \begin{bmatrix} 0.7 & 0.3 & 0 \\ 0.7 & 0 & 0.3 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ for $t = 2, \ldots, 7$.

The results of the procedure testing for two unobservable states from occasions 3 to T-3 are given in Table 5.14. To test for one or two unobservable states, we use the linear combination tests based on the iterative optimisation (stopping criterion: 10^{-9}). The mixture(s) used as proxy bases are chosen randomly

since we showed in Section 5.4 that this choice does not matter in a linear combination framework. As expected, the existing Test M strongly rejects a model structure with no unobservable states (100% of significant results). The linear combination test also rejects a model structure including only one unobservable state at occasion 3 (100% of significant results) but not at occasion 4 (only 5.69% significant results). Finally, the test for two unobservable states presents a Type I error rate close to the expected 5% at occasion 3 whilst there are very few (0.33%) significant results at occasion 4. We can surmise that this is due to fitting an overly complicated model at occasion 4, since statistically at that time, including just one unobservable state would provide an adequate fit of the model. The results obtained from these simulations show that the procedure testing for two unobservable states can work in a capture-recapture setting. But they also show that, due to the complexity of the cell-probabilities constituting the mixtures and bases, in terms of capture-recapture parameters, which can be state and time dependent, it will be hard to predict the circumstances in which the test for one unobservable state will be rejected so that the test for two unobservable states may be applied.

Table 5.14: Test for detecting two unobservable states in a capture-recapture framework, simulation results, percentage of significant test results (5% level),U_{true} denotes the number of unobservable states in the data, U_{fitted} denotes the number of unobservable states used in the fitted and tested model

U_{true}	U_{fitted}	χ^2(%)	CR ($\lambda = \frac{2}{3}$)(%)	G^2(%)	i	*dof*	N	Test
2	0	**100.00**	**100.00**	**100.00**	3	18	300	M
2	0	**100.00**	**100.00**	**100.00**	4	12	300	M
2	1	**100.00**	**100.00**	**100.00**	3	10	299	LC(it)
2	1	5.69	5.69	5.69	4	6	299	LC(it)
2	2	4.68	4.68	4.68	3	4	300	LC(it)
2	2	0.33	0.33	0.33	4	2	300	LC(it)

5.5.3 Testing for one unobservable state in the Canada Geese dataset

When assessing whether the procedure to test for one unobservable state could be used in a capture-recapture setting, we used simulations with extremely large sample sizes. Using the Canada geese dataset presented in Section 3.6 (Chapter 3), we examine whether the procedure could actually be used in real-life conditions. To do so, we set state 3 to be unobservable in the geese dataset (all 3s replaced by 0s) and use our procedure to assess whether the results indicate that one additional state should be defined in the model. Hence, we start from a dataset which seemingly presents only two live states: 1 and 2. The results obtained using the procedure to test for one unobservable state are given in Table 5.15.

Applied to the geese dataset with one unobservable state, Test M yields a significant result at both occasions 3 and 4 whilst the linear combination test for one unobservable state does not, indicating that a model structure including one unobservable state provides an adequate fit to the dataset. Based on these results, an additional state with capture probability 0 would be defined in the model. The procedure works as expected for this example, and this is encouraging in terms of possible applications to capture-recapture datasets.

Table 5.15: Canada geese, state 3 set to unobservable, results of the procedure testing for one unobservable state

U_{true}	U_{fitted}	χ^2(%)	CR ($\lambda = \frac{2}{3}$)(%)	G^2(%)	i	*dof*	Test
1	0	NA	NA	NA	2	NA	M
1	0	**< 0.001**	**< 0.001**	**< 0.001**	3	8	M
1	0	**0.006**	**0.006**	**0.006**	4	4	M
1	1	NA	NA	NA	2	NA	LC(it)
1	1	0.12	0.13	0.13	3	3	LC(it)
1	1	0.22	0.22	0.22	4	1	LC(it)

5.6 Discussion

In this chapter, we have derived general parameter redundancy results for a multinomial mixture and bases model structure including unobservable bases. These results enabled us to adapt the MMLM approach and derive a procedure to test for one or more unobservable bases. We examined different aspects of the procedure such as the structure tested for and the optimisation method used. To sum up, we need $M > U$ mixtures to assess the fit of a model including U unobservable bases: U of the mixtures are used as proxy bases and their cell probabilities estimated separately whilst the remaining mixtures are expressed as linear combinations of both the observable and proxy bases. We showed in this chapter that the tests for unobservable bases should be tests for linear combinations rather than tests for mixtures: they present good distributional properties, any mixtures can be used as proxy bases and more importantly, the remaining mixtures can always be expressed as linear combinations of the mixtures used as proxies and the observable bases. Indeed in some cases, none of the mixtures can be expressed as mixtures of observable and proxy bases, whereas they can be expressed as linear combinations. Table 5.10 was illustrative of this for instance, we also provide here a more extreme example: consider a situation where one of the original mixing probabilities is on a boundary. Mixtures $M1$, $M2$ and $M3$ are generated by bases $B1$, $B2$, $B3$ and $B4$ with respectively, mixing probabilities $\pi_1 = 0.8, \pi_2 = 0, \pi_3 = 0.1, \pi_4 = 0.1$; $\gamma_1 = 0.1, \gamma_2 = 0.2, \gamma_3 = 0.7, \gamma_4 = 0$; $\nu_1 = 0, \nu_2 = 0.1, \nu_3 = 0.3, \nu_4 = 0.6$. The resulting cell-probabilities are displayed in Table 5.16: when B3 and B4 are both unobservable, each of the mixtures presents extreme probabilities in one of the bins; and none of the mixtures can be expressed as a weighted average of the observable bases and other mixtures.

The linear combination tests are based on two steps: parameter estimation and goodness-of-fit assessment. An iterative approach should be used for the

Table 5.16: Mixture-bases cell probabilities for an example with mixing probabilities on the boundary: the extreme probabilities are denoted in red (lowest) and green(highest) for each bin, once bases B3 and B4 are set to unobservable

0.72040	0.05788	**0.00450**	**0.00570**	0.08710	0.08410	0.02052	0.01980	M1
0.09250	0.00836	0.03720	0.16130	**0.56050**	**0.00160**	**0.13424**	0.00430	M2
0.00170	**0.00174**	0.02070	0.08510	0.24050	**0.48050**	0.06176	**0.10800**	M3
0.90000	**0.07200**	0.00500	0.00600	0.00900	0.00500	**0.00100**	**0.00200**	B1
0.00200	0.00300	**0.18000**	**0.80000**	**0.00500**	0.00200	0.00500	0.00300	B2
0.00300	0.00080	0.00100	0.00100	0.79800	0.00100	0.19020	0.00500	B3
0.00100	0.00200	0.00400	0.00800	0.00100	0.80000	0.00700	0.17700	B4

estimation step, so as to avoid local optima, and a stringent enough convergence criterion so that the estimates obtained are not biased (at least 10^{-9}, which seemed to work quite well). The procedure for testing for unobservable bases consists in a step-up approach, performing only one test at each step. The initial step consists of applying the original test for mixtures (i.e. thus testing for 0 unobservable bases); although the hypothesised model structure at this stage is of mixtures and bases, it could be of interest to examine how a test for linear combinations would perform, even at this step. If the original test for mixtures yields a significant result, we test for one unobservable basis by using one of the mixtures as a proxy basis and applying the linear combination test for one unobservable basis. If the result is not significant, we stop here and define an additional basis, which is unobservable, in our model structure. If the test result is significant, we continue to test for two unobservable bases, two mixtures are then chosen to be proxy bases and we use the linear combination test for two unobservable bases and so on and so forth. We have verified, using simulation that the procedure for testing one and two unobservable bases work as expected in theory. Technically, it would be possible to continue this step-up approach to test for U unobservable bases. However, the null hypothesis tested at each step will only be rejected if the unobservable bases are different enough and have specific properties that mean the mixtures cannot be reconstructed without them. Therefore, in practice, it will not be possible to reach a step that tests for many unobservable bases. Note that at each step, only one null hypothesis is tested and this hypothesis is different

for each step. Since "Multiple testing refers to any instance that involves the simultaneous testing of more than one hypothesis" (Romano et al., 2010), we do not believe multiple testing would be an issue in this case.

Finally we have shown that it is possible to apply this procedure to the capture-recapture framework to test for unobservable states. However, the complex nature of capture-recapture parameters and the way they interact to form the relevant cell-probabilities of the mixtures and bases render the properties of corresponding mixtures and bases difficult to assess. Using simulation as well as a real-life dataset, the procedure to test for one unobservable state showed promising results and could be the focus of future research. Indeed it would be of interest to examine the power of the test related to sample size and state-specific properties. Also, recall that the original Test M has a component M.ITEC which is used to assess trap-dependence; thus it would be interesting to investigate how the test for unobservable states might react to this phenomenon. We also showed that a test for two unobservable states can work in theory in a capture-recapture setting, in an extreme situation, with very large sample sizes; but we expect limitations in practice. Indeed the two unobservable states would need to have very specific properties to give rise to mixtures which could not be reconstructed without both of the bases, the animals would also have to have a high probability of moving to and from the unobservable states. It is unclear at this stage what these specific properties should be, and this is an area of future research.

Chapter 6

Modelling individual continuous time-varying covariates using a multi-state framework

6.1 Introduction

In the previous chapters, we presented various types of capture-recapture models for open populations: the CJS model, the multi-state models and finally the more general multievent models. The model parameters (describing survival, capture and transitions) can be constant, time-dependent or state-dependent. They can also be influenced by covariates, which can be external and time-varying (e.g. climate), individual and constant with time (e.g. gender) or individual and time-varying (e.g. weight). The values taken by these covariates can be partially responsible for heterogeneity, they can also be representative of crucial ecological relationships, so that it is important to model them

when appropriate (see for example Pollock, 2002). Individual time-varying covariates are problematic in a capture-recapture context since the covariate value is unknown whenever the animal is not resighted, resulting in an intractable likelihood (Worthington et al., 2015). Several methods have been developed in order to deal with this issue. Nichols et al. (1992) used a coarse discretisation of their individual covariate of interest (body mass), into four categories. These categories were then used as states and the model re-framed as a multi-state model. Amstrup et al. (2005, page 177) note that any continuous covariate could be treated in this way. Bonner and Schwarz (2006) proposed a Bayesian data augmentation method: first modelling the covariate process as well as the relationship between capture-recapture parameters and the covariate; and then using a Bayesian approach based on the complete data likelihood to estimate the parameters. Catchpole et al. (2008) introduced the trinomial method, which only takes into account the events associated with observed (i.e. known) values of the covariate, thus using a conditional likelihood approach. More recently, Langrock and King (2013) used hidden Markov model machinery, based on a state-space formulation of the original model. They modelled the covariate process and finely discretised the covariate space in order to use an approximate likelihood approach, based on a sum instead of an integral, thus using all of the available data, unlike the trinomial method. Finally, Worthington et al. (2015) presented a two-step method: multiple imputations of the missing covariate values assuming they are missing at random, based on a covariate process model fitted to the observed data, followed by the estimation of the demographic parameters resulting from the likelihood conditional on observed and imputed values of the covariate.

In this chapter, we focus on the multi-state approach proposed by Nichols et al. (1992). This method has been criticised, mainly on the grounds of the coarse discretisation used, which results from an arbitrary decision and may lead to losing valuable information (Worthington et al., 2015; Langrock and

King, 2013; Bonner et al., 2010), or to a violation of the assumption of same behaviour for animals in the same state (Bonner and Schwarz, 2006). However, the multi-state approach possesses the advantage of not needing to model the covariate process since the parameters can be estimated freely as state-dependent (Bonner and Schwarz, 2006), and, most importantly, missing values due to non-captures of animals are automatically treated in this framework. Since the main criticism pertains to the coarse discretisation of the covariate rather than the use of multi-state models per se, we investigate the performance of a multi-state approach using a fine discretisation of the covariate space, in a simple CJS capture-recapture context. In this chapter, we focus on the initial goal of investigating whether the multi-state model produces good results rather than assessing its performance relative to all the other methods currently developed. In Section 6.2, we first examine a situation where survival is strongly dependent on body-mass using simulation. We then apply this methodology to a dataset of great crested newts (*Triturus cristatus*) to explore a potential link between body mass and demographic parameters (Section 6.3.1). In Section 6.3.2, we revisit the North American meadow vole (*Microtus pennsylvanicus*) dataset from Nichols et al. (1992) to assess whether a finer discretisation affects the results previously obtained. Finally, in Section 6.4, using simulation, we briefly explore the performance of this approach for possible extensions such as incorporating measurement error or accommodating missing data (i.e. missing covariate value when the animal is resighted). We conclude in Section 6.5. All the multi-state models fitted in this chapter were fitted using Matlab.

6.2 Simulation study

6.2.1 Fine discretisation and multi-state model parametrisation

In order to finely discretise the covariate space, we follow the approach of Langrock and King (2013): defining the essential range as $0.8w_{min}$ to $1.2w_{max}$, with w_{min} and w_{max} respectively denoting the minimum and maximum of the observed covariate values; then dividing this range into m equal intervals. The boundaries of these intervals will represent the "live" states and are relabelled accordingly from 1 to $m+1$, with the animals assigned to the boundary closest to their covariate value. Taking the great crested newts' body mass as an example, using $m = 20$, $w_{min} = 4.5$g and $w_{max} = 21.7$g, this leads to the following boundaries:{4.0, 5.1, 6.2, 7.3, 8.4, 9.5, 10.6, 11.7, 12.8, 13.9, 15.0, 16.1, 17.2, 18.3, 19.4, 20.5, 21.6, 22.7, 23.8, 24.9, 26.0}. An illustration of conversion from example weights to a multi-state format is given in Table 6.1. For example animals with weights of 8.3 and 8.5 are both assigned to state 5 since they are closest to the boundary 8.4.

Bonner and Schwarz (2006) observed that using a large number of states can result in identifiability issues due to the considerable number of parameters to estimate. Keeping this in mind, we simplify the parametrisation usually associated with state-dependence in classic multi-state models. For the survival probability in state r, denoted by $\phi^{(r)}$, instead of using one parameter

Table 6.1: Example of conversion from original weightings to a multi-state format, NA denotes missing values corresponding to occasions on which the animals was not captured

Occasion	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**
Original weight	8.5	8.3	NA	10.3	12.9	13.0	16.3	11.4
Multi-state format	5	5	0	7	9	9	12	8

per state, we model the dependence on weight using a linear logistic model (North and Morgan, 1979), as is usually done with continuous covariates: $\text{logit}(\phi^{(r)}) = \beta_0 + \beta_1 r$, with $\text{logit}(x) = \log(x/(1-x))$ The capture probability, set to be constant over time and states, and denoted by p, is modelled using a logit link: $\text{logit}(p) = \beta_2$. We also choose a simple structure for the transition probabilities, which considerably reduces the number of parameters to estimate. In this case, we use multinomial logistic-type links (in order for the transition matrix to remain row-stochastic), one for weight increase and the other for weight decrease, with the transition probabilities being a function of the difference between arrival state and departure state (i.e. the weight jump). Additionally, we constrain the transition probabilities of large jumps to be zero, using as threshold the maximum observed weight jump between two consecutive occasions plus $0.1m$. Using the probability of remaining in the same state as the reference level, and using r and s to respectively denote the arrival and departure states, if $|r-s| \leq 0.1m$,

- for $r < s$ (weight increase)

$$\psi_{r,s} = \frac{\exp(\beta_3 + |r-s|\,\beta_4)}{1 + \sum_{s>r,s=1}^{m+1} \exp(\beta_3 + |r-s|\,\beta_4) + \sum_{s<r,s=1}^{m+1}, \exp(\beta_5 + |r-s|\,\beta_6)}$$

- for $r > s$ (weight decrease)

$$\psi_{r,s} = \frac{\exp(\beta_5 + |r-s|\,\beta_6)}{1 + \sum_{s>r,s=1}^{m+1} \exp(\beta_3 + |r-s|\,\beta_4) + \sum_{s<r,s=1}^{m+1}, \exp(\beta_5 + |r-s|\,\beta_6)}$$

- and $\psi_{r,r} = 1 - \sum_{s\neq r,s=1}^{m+1} \psi_{r,s}$

Hence, instead of estimating $(m+1)(m+1) - m$ parameters, we estimate only four parameters: β_3, β_4, β_5 and β_6. The probability of remaining in the same state is the reference level and the parameters are interpreted relative to this probability. Although the parametrisation is simple, an interpretation

of the estimates at first glance is not as straightforward, as illustrated using a simple example (note also that this is not a "real" multinomial regression with independent predictors). Recall that another way of expressing the model for a weight increase, for instance, would be: $\log(\psi^{r,s}/\psi^{r,r}) = \beta_3 + |r-s|\,\beta_4$. Let us suppose that the MLEs obtained were $\hat{\beta}_3 = 0.64$, $\hat{\beta}_4 = -0.25$, $\hat{\beta}_5 = 0.42$ and $\hat{\beta}_6 = -0.58$. Then, the log-odds of increasing state relative to remaining in the same state decreases by 0.25 unit when the state value increases by one unit. Using numerical examples, the log-odds of increasing the state value by 1, 2 or 3 would be respectively: $0.64 - 0.25 = 0.39$, $0.64 - 0.25 \times 2 = 0.14$, and $0.64 - 0.25 \times 3 = -0.11$. The same reasoning applies to a weight decrease: the log-odds of decreasing state relative to remaining in the same state decreases by 0.58 unit when the state value decreases by one unit. Using numerical examples, the log-odds of decreasing state by respectively, 1, 2 or 3 compared to remaining the same weight is $0.42 - 0.58 = -0.16$, $0.42 - 2 \times 0.58 = -0.74$ and $0.42 - 3 \times 0.58 = -1.32$. In other words, for this example, the animals would be more likely to increase by 1 or 2 state-units than to remain in the same weight category, but would be less likely to decrease in weight or to increase by more than 2 state-units.

The choice of this structure for transition probabilities was based on basic intuitive considerations regarding weight: that the animals could increase or decrease weight in different manners and that the probabilities of changing weight would depend on the change in weight value between times.

6.2.2 Simulation results

We used simulation in order to evaluate how well the multi-state approach performed to estimate the survival parameters as well as the capture probability, in the presence of a strong link between survival and weight. 250 datasets of 500 individuals, arising from a simple CJS model, were simulated, with a

strong logistic-type link between the survival probability of animal k from t to $t+1$ ($\phi_{k,t}$) and a simulated individual continuous covariate representing body mass $w_{k,t}$: $\text{logit}(\phi_{k,t}) = \beta_0 + \beta_1 w_{k,t}$, with $\beta_0 = -5.05$ and $\beta_1 = 0.7$.

The covariate values were simulated in R, using the homogeneous order 1 autoregressive process, proposed by Langrock and King (2013),

$$w_{k,t} = \eta(w_{k,t-1} - \mu) + \mu + \sigma\epsilon_t \text{ with } \epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0,1);$$

setting the values to $\sigma = 1.38$, $\eta = 0.81$, $\mu = 18.6$. The initial weights were simulated from a normal distribution with mean 7.5 and standard deviation 2.48. All these covariate process values are roughly based on the estimates obtained when fitting this model to the great crested newts dataset (the values were slightly modified in order to allow for a stronger increase in weight over time). The evolution of weight over time is illustrated in Figure 6.1 and the link between survival and weight in Figure 6.2.

Varying levels of capture probabilities were examined: $p = 0.9$, 0.6 and 0.3, as well as increasing levels of fine discretisation: $m = 10$, 20, 40 and 60. The results obtained by fitting a multi-state model with 7 parameters, as described in subsection 6.2.1, are presented in Table 6.2 in terms of mean relative bias with 2.5 and 97.5 percentiles, mean of the 95% confidence interval widths of the parameter estimates and coverage probability of these confidence intervals. The relative bias is equal to $(\hat{\beta} - \beta)/\beta$, with $\hat{\beta}$ denoting the estimated parameter value and β the true value. When using the multi-state framework, the link between survival and weight is modelled as a function of the state r: $\text{logit}(\phi^{(r)}) = \beta_0 + \beta_1 r$. Therefore the estimates obtained cannot be directly compared to the true values; they need to be rescaled back to the original weight scale. The state value r is obtained from the discretised weight boundary wd, the minimum of the essential range min and the length of the equal intervals between boundaries, denoted by $length$: $r = (\text{wd} - \text{min})/(\text{length}) + 1$.

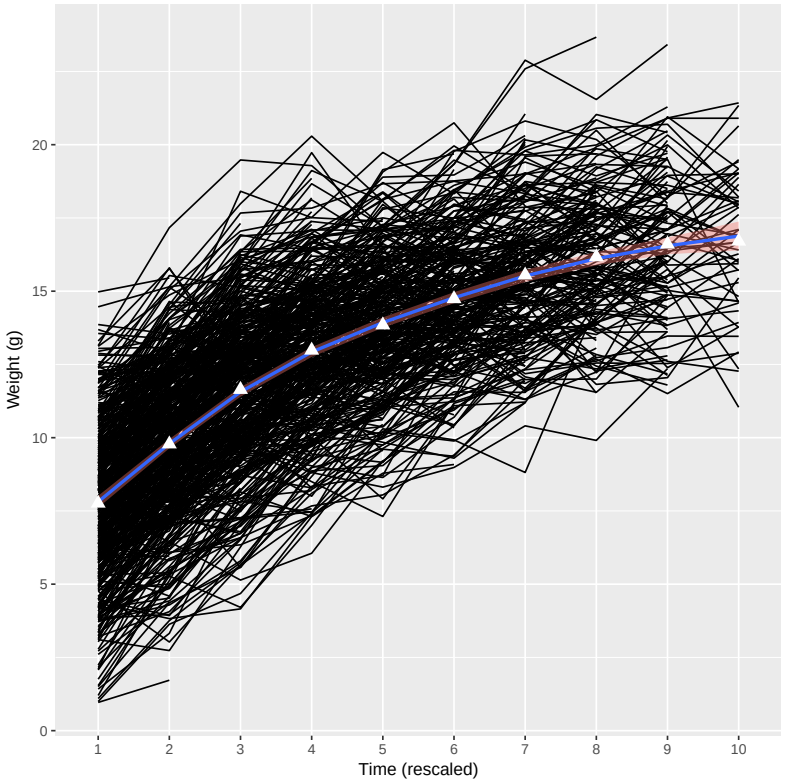


Figure 6.1: Weight evolution over time for a simulated dataset with 10 release occasions: individual profiles and average trend. The mean is represented by triangles and the standard error by the shaded area. Time has been rescaled so that 1 represents the initial release occasion of the animal, capture and survival probabilities were set to 1 in order to illustrate only the covariate process.

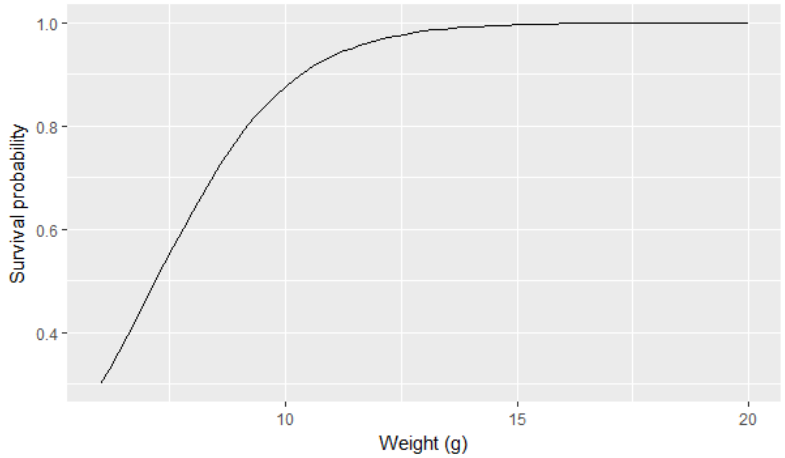


Figure 6.2: Illustration of the link between survival and weight chosen for the simulations

In order to distinguish parameters on a state scale from those on a weight scale, we denote the multi-state parameters by the superscript MS.

$$\beta_0^{MS} + r\beta_1^{MS} = \beta_0^{MS} + \left(\frac{\text{wd} - \min}{\text{length}} + 1\right)\beta_1^{MS}$$

By simply re-arranging the terms, we obtain

$$\beta_0^{MS} + r\beta_1^{MS} = \beta_0^{MS} + \beta_1^{MS} - \beta_1^{MS}\frac{\min}{\text{length}} + \frac{\beta_1^{MS}}{\text{length}}\text{wd}$$

Hence the parameters on the original weight scale are

$$\beta_0 = \beta_0^{MS} + \beta_1^{MS} - \beta_1^{MS}\frac{\min}{\text{length}}$$

$$\text{and } \beta_1 = \frac{\beta_1^{MS}}{\text{length}}.$$

There will obviously be a slight loss in precision due to the fact that the rescaling from multi-state to weight goes back to the *discretised* weights, not the original weights. The standard errors of the parameter estimates are obtained using the square root of the diagonal terms of the inverse Hessian numerically evaluated at the MLE. We also need to transform these standard errors obtained to compute the confidence intervals on the weight scale, using the classic formula of variance for a sum of variables: $\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$. It follows that

$$\text{Var}(\hat{\beta}_0) = \text{Var}(\hat{\beta}_0^{MS}) + \text{Var}(\hat{\beta}_1^{MS})\left[1 - \frac{\min}{\text{length}}\right]^2 + 2\left[1 - \frac{\min}{\text{length}}\right]\text{Cov}(\hat{\beta}_0^{MS}, \hat{\beta}_1^{MS})$$

$$\text{Var}(\hat{\beta}_1) = \left(\frac{1}{\text{length}}\right)^2 \text{Var}(\hat{\beta}_1^{MS})\,.$$

Finally, for the capture probability, p, the standard errors on the probability scale are derived using the delta method, with ilogit denoting the inverse logit link: $\text{ilogit}(x) = \exp(x)/(1 + \exp(x))$ and $\text{se}(\hat{p}) = \text{ilogit}(\hat{\beta}_2)(1 -$

ilogit($\hat{\beta}_2$))se($\hat{\beta}_2$) (see for example Cooch and White, 2014, p.55, Chapter 4).

The simulation results show that a multi-state approach with fine discretisation produces very good results. Indeed over all levels of capture probabilities examined, $m = 10$ produces biased estimates (around 10% for the intercept). However, the biases are drastically reduced for $m = 20$: around 1 % for $p = 0.9$, 2% for $p = 0.6$ and 3% for $p = 0.3$. We note that higher discretisations of $m = 40$ and $m = 60$ do not seem to change the results much: the intercept is estimated with no bias, but the relative bias of the weight coefficient are slightly increased, although they remain low (e.g. 5 % for $p = 0.3$). Finally we note that the relative bias of the capture probability estimates increases slightly as the true capture probability decreases and is constant over all levels of discretisation: unbiased for p=0.9 to a 1% relative bias for p=0.6 and 3% for p=0.3. The confidence interval widths do not seem to change much across the discretisations, however they become larger as the capture probability decreases, as expected. Finally the coverage probabilities of the confidence intervals are very high for $m \geq 20$, around 96% for the survival parameters, but they are lower for the capture probability around 90% for $p = 0.6$ and $p = 0.3$. On the other hand, the confidence interval widths are extremely narrow for the capture probability estimates. Based on these simulations, a multi-state approach for modelling time-varying individual covariates seems to work well for $m \geq 20$, even for a low capture probability and despite the use of a very simple structure for the transition probabilities. Based on these simulation results, $m = 20$ presents good properties in terms of computational time and estimation performance, whereas $m = 40$ and $m = 60$ took much longer to run for no significant gain in estimate precision (see Table 6.2). As an aside, the number of available simulations is consistently lower than 250 for $m = 10$ because the obtained estimates were associated with negative variance estimates for some datasets and these were not taken into account. Since this did not happen for $m \geq 20$ (apart for $p = 0.3$, for only 7 datasets), we did not

deem it necessary to investigate the optimisation problem further.

Seeing that the multi-state approach with fine discretisation is shown to work well, the next step to investigate in future research would be to assess its relative performance compared to the existing methods described in Section 6.1.

6.3 Applications

6.3.1 Great crested newts

We use a CJS-format dataset of great crested newts, captured on a field study site at the University of Kent campus, using funnel traps, from 2003 to 2014 and identified using their unique belly patterns. The dataset consists of 108 individuals: 50 females and 58 males, whose weight information was collected on capture. The captures occurred from around March until around mid-July, when no newt was captured (Griffiths et al., 2010). The data we used were in an annual format: coded as 1 if a newt was captured at least once during the capture season and 0 if not; annual weight data was also available for these newts.

We use a multi-state framework to investigate a potential effect of weight on the survival or capture probability. Recall that our simulations from Section 6.2.2 were roughly based on the weight process used by Langrock and King (2013), fitted to the great crested newts. Since our results showed that a discretisation in $m = 20$ intervals presented good properties, we set m to 20 for our application.

First, out of curiosity, we compared the results obtained by fitting a simple model using the Langrock and King (2013) HMM approach and the multi-state approach presented in this chapter (setting $m = 20$). We fitted a model with constant capture probability, weight/state-dependent survival probability

Table 6.2: Modelling weight via a multi-state approach using an increasingly fine discretisation and different level of capture probabilities: simulation results; RB denotes the relative bias, q the quantiles, CIW the confidence interval widths, CP the coverage probability and N the number of usable datasets (i.e. omitting the datasets presenting negative variance estimates).

p_{true}	m	Parameter	RB($q_{0.025}$, $q_{0.975}$)	CIW	CP	N
		β_0	-0.09 (-0.22, 0.05)	1.37	0.74	
	10	β_1	-0.07 (-0.18 ,0.05)	0.16	0.76	220
		p	0.00 (-0.02 , 0.01)	0.04	0.94	
		β_0	-0.01 (-0.13,0.15)	1.47	0.96	
	20	β_1	-0.01 (-0.11, 0.13)	0.18	0.97	250
		p	0.00 (-0.02,0.02)	0.04	0.94	
0.9						
		β_0	0.01 (-0.12,0.17)	1.49	0.96	
	40	β_1	0.01 (-0.10,0.15)	0.18	0.96	250
		p	0.00 (-0.02,0.02)	0.04	0.94	
		β_0	0.01(-0.12,0.17)	1.5	0.96	
	60	β_1	0.01 (-0.09,0.15)	0.18	0.96	250
		p	0.00 (-0.02,0.02)	0.04	0.94	
		β_0	-0.09 (-0.24, 0.05)	1.65	0.75	
	10	β_1	-0.06 (-0.17, 0.08)	0.20	0.86	163
		p	-0.02 (-0.06, 0.03)	0.06	0.92	
		β_0	-0.02 (-0.18, 0.16)	1.76	0.94	
	20	β_1	0.00 (-0.12, 0.16)	0.22	0.98	250
		p	-0.02 (-0.07, 0.03)	0.06	0.90	
0.6						
		β_0	0.00 (-0.15 , 0.17)	1.79	0.96	
	40	β_1	0.02 (-0.11, 0.18)	0.22	0.96	250
		p	-0.01 (-0.07, 0.03)	0.06	0.90	
		β_0	0.00 (-0.15 0.18)	1.79	0.96	
	60	β_1	0.02 (-0.10, 0.17)	0.22	0.98	250
		p	-0.01 (-0.07, 0.03)	0.06	0.90	
		β_0	-0.10 (-0.32, 0.12)	2.29	0.82	
	10	β_1	-0.03 (-0.22, 0.16)	0.30	0.92	119
		p	-0.03 (-0.12, 0.07)	0.06	0.91	
		β_0	-0.02 (-0.20, 0.19)	2.41	0.96	
	20	β_1	0.03 (-0.14, 0.23)	0.31	0.98	243
		p	-0.03 (-0.12, 0.06)	0.06	0.91	
0.3						
		β_0	0.00 (-0.20, 0.25)	2.46	0.96	
	40	β_1	0.05 (-0.13, 0.26)	0.31	0.97	250
		p	-0.03 (-0.12, 0.07)	0.06	0.91	
		β_0	0.00 (-0.19, 0.25)	2.46	0.96	
	60	β_1	0.05 (-0.12, 0.27)	0.31	0.97	250
		p	-0.03 (-0.12, 0.07)	0.06	0.91	

through a logistic link, and transitions structured similarly to Section 6.2.1. Table 6.3 presents the common parameters estimated by both models and shows that both approaches yield very similar results (only the intercept estimate for survival is slightly different); they both lead to the same estimate and conclusion regarding weight, that it has no significant effect on survival.

Then, using the multi-state approach with $m = 20$, in order to further investigate whether weight intervened once other effects had been accounted for, we performed some model selection based on a step-up procedure starting with a model defined by constant capture, survival and transition probabilities, which is termed the level 0 model (McCrea and Morgan, 2011). In the next step, models are fitted by allowing just one level of dependence for each of the parameters and this is done for all the possible types of dependence for each of the parameters; these form the level 1 models. For instance a model with time-dependent survival, constant capture and constant transition probabilities is a level 1 model. The best model from level 1 is chosen using the AIC_c as the criterion to select the best model at each stage. It is then used as the starting point for the next level models, where one level of dependence will be added for each of the parameters. The best model from this level then serves as the starting point for the next level and so forth. The AIC_c was computed according to the formula given below: $\log(L(\hat{\theta}))$ denotes the log-likelihood evaluated at its maximum, d denotes the number of estimable parameters of

Table 6.3: Great crested newts: capture and survival estimates and associated standard errors; multi-state (MS) and HMM approaches. Note that the conversion from multi-state results to original weight scale only concern the weight-dependent survival parameters, the capture probability is not expressed in relation to weight and thus is not expressed on the weight scale.

	MS	MS (weight scale)	HMM
ϕ, intercept: β_0	0.91 (0.569)	0.67 (0.745)	0.63 (0.763)
ϕ, weight coefficient: β_1	0.09 (0.069)	0.08 (0.063)	0.08 (0.064)
logit(p)	1.55 (0.188)	-	1.55 (0.188)

the model and n the effective sample size (Burnham and Anderson, 2002, page 66).

$$AIC_c = -2\log\{(L(\hat{\theta})\} + 2d + \frac{2d(d+1)}{n-d-1}$$

When the sample size is large relative to the number of estimable parameters, the AIC_c will be approximately equal to the AIC: $-2\log\{L(\hat{\theta})\} + 2d$. We note that the definition of effective sample size in the capture-recapture framework does not seem straightforward due to the multiple measurement points for each animal, and there does not seem to be much literature on the subject. Cooch and White (2014, Chapter 19) state that "*The appropriate value for effective sample size in mark-recapture models remains an unresolved issue and more research on the topic is warranted*". Burnham and Anderson (2002, page 332) suggest using the number of animals captured at least once, or the number of potential recaptures. Programs commonly used for capture-recapture analyses such as MARK or E-SURGE also differ in their definition of the effective sample size, with MARK using the total number of captures and recaptures, leaving out the last year (see response from Cooch on the forum `http://www.phidot.org/forum/viewtopic.php?f=1&t=3136`) and E-SURGE using the total number of captures and recaptures minus the number of individuals removed or censored (see response from Choquet on the forum `http://www.phidot.org/forum/viewtopic.php?f=5&t=1597`). We chose to use the sample size defined in program E-SURGE, simply because we have often used this program to check our results from the simpler models. Since there were no individuals removed, the effective sample was 333 (this would be 288 with the MARK definition or 225 using the number of potential recaptures).

At each stage/level of the step-up procedure, we select as best model the one with the smallest AIC_c. We use the rough rule of thumb recommended in Burnham and Anderson (2004, p.70), that $\Delta(AIC_c) <= 2$ indicates strong support for the model, with $\Delta(AIC_c)$ being the difference between the AIC_c

value of a model considered within a given level and the AIC_c value of the best model within the same level. Although the discretisation was set to $m = 20$, the number of states was actually 17 in the multi-state dataset, since some of the larger weight categories were not represented. Throughout the model-fitting process, we encountered the issue of very large or negative variance estimates at the maximum likelihood. Upon further examination, these were due either to convergence to a local optimum rather than the global one, or to parameters being estimated at a boundary. As a safeguard against local minima, we checked that the deviances obtained for all models fitted at level L were smaller than that of the best model selected at level $L - 1$, which is nested within the more complicated models of level L. Thus, we re-ran the optimisation process when necessary. Following the recommendations of McCrea and Morgan (2011), we also used the maximum-likelihood estimates obtained from the best model at level $L - 1$ as initial values for the optimisation for models at level L. Finally, the boundary estimates we encountered were probabilities estimated as 1, so we fixed these values and estimated only the remaining parameters. After making these corrections, we no longer obtained aberrant variance estimates for the fitted models. The results of the model fitting performed on the great crested newts dataset are presented in Table 6.4, with the best model at each level indicated in red. The models are denoted based on the GEMACO notation used in program E-SURGE: p, ϕ and ψ respectively denote the capture, survival and transition probabilities; (.) denotes a constant parameter, t time-dependence, g gender effect, w weight effect (modelled through a logistic regression link as described in section 6.2), '+' indicates additive effects and '$*$' interactive effects.

Note that categorical effects are modelled using dummy variables (one variable per category, no intercept); and additive categorical effects involving categorical variables will result in multicollinearity. Indeed, for each variable, one category will be associated to a value of 1; hence the sum of the dummy vari-

ables associated with each categorical effect will sum to 1 for all possible combinations. Take for instance an additive gender and time effect characterised by the following indicators: male = 0, female = 1, $t_1 = 1$, $t_2, \ldots, t_{11} = 0$, it is obvious that the gender indicators and the time indicators will each sum to 1, whatever is the combination of indicators. To solve this problem, we followed the method given in Choquet and Nogue (2006) and dropped the dummy variable associated to one of the categories: for example, with a parametrisation $g + t$, parameters were estimated from $t = 2$ onwards. For the transition probabilities, D denotes the departure state (i.e. state at time $t - 1$) and J the size of the jump between arrival state (i.e. state at time t) and departure state, $(D + J)_{diff}$ denotes that the link parameters are different for a weight increase than for a weight decrease as in the model described in Section 6.2.1, except that we did not constrain any of the transition probabilities to 0. The notation (B) indicates that parameters estimated at a boundary have been fixed and the number of estimable parameters reduced consequently.

Since our simulation results as well as the first step of simple model fitting to the newt dataset both showed that accurate results were obtained when using a simple structure for the transition probabilities, we only considered a few simple structures for these in our model fitting. Our goal is not to model the transitions as accurately as possible, but to use the multi-state framework to fit models with weight-dependent capture and/or survival probabilities, and accurately estimate the potential weight-effect.

We note that we did not fit the interactive model $p(g), \phi(.), \psi\{(D+J)_{diff} * g * t\}$ at level 4, due to the considerable number of parameters involved (135), with 132 for the transition structure (in this case it would make more sense to actually model the covariate process rather than using an over-parametrised possibly inaccurate structure for the transition process). We stopped the model fitting at level 4 since weight-dependence had not entered the model for survival nor for capture by that point.

Table 6.4: Newts, model fitting results, the best model at each level is denoted in red. The last model from the list corresponds to the best model from level 4, including environmental covariates

Level	Model	Deviance	d	AIC	AIC_c
0	$p(.), \phi(.), \psi(.)$	1613.14	3	1619.14	1619.21
1	$p(.), \phi(.), \psi(D+J)_{diff}$	1225.56	8	1241.56	1242.01
	$p(.), \phi(.), \psi(D+J)$	1311.15	5	1321.15	1321.33
	$p(.), \phi(.), \psi(J)$	1313.60	4	1321.60	1321.72
	$p(w), \phi(.), \psi(.)$	1531.53	4	1539.53	1539.65
	$p(g), \phi(.), \psi(.)$	1583.11	4	1591.11	1591.23
	$p(.), \phi(.), \psi(g)$	1608.59	4	1616.59	1616.71
	$p(.), \phi(.), \psi(D)$	1609.84	4	1617.84	1617.96
	$p(t), \phi(.), \psi(.)$	1592.83	13	1618.83	1619.97
	$p(.), \phi(w), \psi(.)$	1611.50	4	1619.50	1619.62
	$p(.), \phi(t), \psi(.)$	1593.03	13	1619.03	1620.17
	$p(.), \phi(g), \psi(.)$	1612.98	4	1620.98	1621.10
	$p(.), \phi(.), \psi(t)$	1597.39	13	1623.39	1624.53
2	$p(.), \phi(.), \psi\{(D+J)_{diff} * g\}$	1197.78	14	1225.78	1227.10
	$p(g), \phi(.), \psi\{(D+J)_{diff}\}$	1214.06	9	1232.06	1232.62
	$p(.), \phi(.), \psi\{(D+J)_{diff} + g\}$	1216.49	9	1234.49	1235.05
	$p(w), \phi(.), \psi(D+J)_{diff}$	1221.20	9	1239.20	1239.76
	$p(.), \phi(w), \psi(D+J)_{diff}$	1223.89	9	1241.89	1242.45
	$p(t), \phi(.), \psi\{D+J)_{diff}$	1205.25	18	1241.25	1243.43
	$p(w), \phi(t), \psi(D+J)_{diff}$	1205.45	18	1241.45	1243.63
	$p(.), \phi(.), \psi\{(D+J)_{diff} + t\}$	1210.65	18	1246.65	1248.83
	$p(.), \phi(g), \psi(D+J)_{diff}$	1225.40	9	1243.40	1243.95
	$p(.), \phi(.), \psi\{(D+J)_{diff} * t\}$	1128.15	68	1264.15	1299.69
3	$p(g), \phi(.), \psi\{(D+J)_{diff} * g\}$	1186.28	15	1216.28	1217.80
	$p(w), \phi(.), \psi\{(D+J)_{diff} * g\}$	1191.99	15	1221.99	1223.51
	$p(.), \phi(w), \psi\{(D+J)_{diff} * g\}$	1196.09	15	1226.09	1227.61
	$p(t), \phi(.), \psi\{(D+J)_{diff} * g\}$	1177.47	24	1225.47	1229.37
	$p(.), \phi(t), \psi\{(D+J)_{diff} * g\}$	1177.67	24	1225.67	1229.57
	$p(.), \phi(g), \psi\{(D+J)_{diff} * g\}$	1197.61	15	1227.61	1229.13
	$p(.), \phi(.), \psi\{(D+J)_{diff} * (g+t)\}$	1184.99	24	1232.99	1236.88
	$p(.), \phi(.), \psi\{(D+J)_{diff} * (g*t)\}$	1060.83	134	1328.83	1511.55
4	$p(g+t), \phi(.), \psi\{(D+J)_{diff} * g\}(B) : p_2 = p_6 = 1$	1166.11	23	1212.11	1215.68
	$p(g), \phi(w), \psi\{(D+J)_{diff} * g\}$	1184.16	16	1216.16	1217.88
	$p(g), \phi(t), \psi\{(D+J)_{diff} * g\}(B) : \phi_3 = 1$	1166.05	24	1214.05	1217.94
	$p(g+w), \phi(.), \psi\{(D+J)_{diff} * g\}$	1186.03	16	1218.03	1219.75
	$p(g), \phi(g), \psi\{(D+J)_{diff} * g\}$	1186.23	16	1218.23	1219.95
	$p(g.w), \phi(.), \psi\{(D+J)_{diff} * g\}$	1184.29	17	1218.29	1220.24
	$p(g), \phi(.), \psi\{(D+J)_{diff} * g + t\}$	1173.50	25	1223.49	1227.72
	$p(g*t), \phi(.), \psi\{(D+J)_{diff} * g\}$	1183.62	35	1253.62	1262.11
	$p(g+SR), \phi(NAR+WT+NAR*WT),$ $\psi\{(D+J)_{diff} * g\}$	1174.64	19	1212.64	1215.07

At level 4, according to the AIC_c criterion the best model, with AIC_c=1215.68, is $p(g+t), \phi(.), \psi\{(D+J)_{diff} * g\}$ with boundary parameters p_2 and p_6 set to 1 for both males and females. Since $p(g), \phi(w), \psi\{(D+J)_{diff} * g\}$ presents a $\Delta(AIC_c)$=2.20, we use a likelihood ratio test to determine the significance of the weight effect by comparing this model with the nested model $p(g), \phi(.), \psi\{(D + J)_{diff} * g\}$ (the best model from level 3). The test-statistic is defined as$-2\{\log(L_{M0}) - \log(L_{M1})\}$ with $\log(L_{M0})$ the maximised log-likelihood value for the simpler nested model denoted by $M0$ and $\log(L_{M1})$ the maximised log-likelihood value for the more complex model denoted by $M1$; this test-statistic follows a chi-square distribution with degrees of freedom equal to the number of additional parameters estimated in M1 (see for example McCrea and Morgan, 2014, p. 16). The likelihood-ratio test shows that the weight effect is non-significant (LRT-statistic = 2.12, $p = 0.14$ for 1 d.o.f.); alternatively we can draw the same conclusion from the weight coefficient estimate obtained for this model: 0.10 with 95% confidence interval (-0.04, 0.24).

Finally, since time-dependence intervenes in the best level 4 model for capture probabilities, that $\Delta(AIC_c) = 2.26$ for model $p(g), \phi(t), \psi\{(D+J)_{diff} * g\}$ with ϕ_3 fixed to 1, and that climatic data was shown to influence demographic parameters in previous analyses (Griffiths et al., 2010), we also fit a model with time-varying environmental covariates for both capture and survival probabilities, the results of this model are presented in Table 6.5.

For the capture probabilities, we considered the spring rainfall SR (rainfall over March, April and May for each calendar year), since it corresponds to the period where newts are in the ponds (Griffiths et al., 2010). We chose the mean minimum winter temperature WT and the non-aquatic rainfall NAR (thus termed since it is the rainfall over the period where newts are out of the water) for survival as they were shown to be important predictors in the previous analyses conducted by Griffiths et al. (2010). We computed the NAR

as the rainfall over the period between capture seasons (i.e. rainfall from June to December over year y and January-February in year $y+1$). All the environmental covariates were standardised, with missing data replaced by 0 in this format (which is equivalent to a replacement by the overall mean). Table 6.5 presents the results for the competing best models (model including environmental covariates and best model from level 4), in terms of parameter estimates and associated 95% confidence intervals.

All models have the same parametrisation for the transition probabilities and, as expected, yield the same results for all those parameters. They are presented more for informative purposes since, as illustrated by the Section 6.2.1 example, their interpretation is not intuitive. For males, the log-odds of decreasing state value (i.e. weight category) relative to remaining in the same weight category increases with the departure state value (for a jump size held constant) whilst this log-odds decreases with the jump size (for a departure state held constant). The log-odds of increasing weight category relative to remaining in the same weight category decreases with the departure state value (for a jump size held constant) as well as with the jump size (for a departure state held constant). For females, the log-odds of decreasing state value (i.e. weight category) relative to remaining in the same weight category does not vary significantly with the departure state value (for a jump size held constant) whilst this log-odds decreases with the jump size (for a departure state held constant). The log-odds of increasing weight category relative to remaining in the same weight category decreases with the departure state value (for a jump size held constant)and it also decreases with the jump size (for a departure state held constant). Note also that we are cautious regarding the interpretation of the transition parameters since, as mentioned previously in this section, the transition process was not the focus of our modelling.

The environmental model (denoted by E in Table 6.5) show that the capture probability significantly decreases when spring rainfall increases and that

Table 6.5: Newts, best model results, the significant effects are denoted in bold; E denotes the model with environmental covariates, L4 the best model from Level 4. For the transition parameters, I and D respectively denote weight increase and weight decrease, r denotes the departure state.

Parameter	E		L4	
	Estimate	95% CI	Estimate	95% CI
ϕ: intercept	1.79	(1.32 , 2.25)	1.65	(1.32 , 1.98)
ϕ: β_{WT}	**-0.59**	**(-1.10 , -0.09)**	-	-
ϕ: β_{NAR}	-0.87	(-2.07 , 0.32)	-	-
ϕ: β_{WT*NAR}	0.00	(-1.00 , 1.00)	-	-
p_{males}: intercept	**2.29**	**(1.64 , 2.94)**	1.90	(0.33 , 3.47)
$p_{females}$: intercept	**0.98**	**(0.50 , 1.47)**	0.56	(-1.05 , 2.17)
p: β_{SR}	**-0.41**	**(-0.79 , -0.03)**	-	-
p: β_{t_3}	-	-	0.94	(-1.62 , 3.50)
p: β_{t_4}	-	-	-0.42	(-2.76 , 1.92)
p: β_{t_5}	-	-	1.65	(-0.85 , 4.16)
p: β_{t_7}	-	-	0.33	(-1.54 , 2.21)
p: β_{t_8}	-	-	1.10	(-0.85 , 3.06)
p: β_{t_9}	-	-	-0.71	(-2.39 , 0.97)
p: $\beta_{t_{10}}$	-	-	0.79	(-1.07 , 2.65)
p: $\beta_{t_{11}}$	-	-	0.14	(-1.69 , 1.97)
ψ_{males}, D:intercept	-4.32	(-7.57 , -1.07)	-4.32	(-7.57 , -1.07)
ψ_{males}, D:β_r	**0.45**	**(0.10 , 0.81)**	**0.45**	**(0.10 , 0.81)**
ψ_{males}, D:β_{jump}	**-1.05**	**(-1.53 , -0.57)**	**-1.05**	**(-1.53 , -0.57)**
ψ_{males}, I:intercept	3.58	(1.78 , 5.38)	3.58	(1.78 , 5.38)
ψ_{males}, I:β_r	**-0.34**	**(-0.57 , -0.11)**	**-0.34**	**(-0.57 , -0.11)**
ψ_{males}, I:β_{jump}	**-1.13**	**(-1.42 , -0.85)**	**-1.13**	**(-1.42 , -0.85)**
$\psi_{females}, D$:intercept	-0.39	(-4.78 , 4.01)	-0.39	(-4.78 , 4.01)
$\psi_{females}, D$:βr	0.09	(-0.29 , 0.48)	0.09	(-0.29 , 0.48)
$\psi_{females}, D$:β_{jump}	**-1.40**	**(-2.34 , -0.46)**	**-1.40**	**(-2.34 , -0.46)**
$\psi_{females}, I$:intercept	3.99	(1.05 , 6.94)	3.99	(1.05 , 6.94)
$\psi_{females}, I$:β_r	**-0.30**	**(-0.58 , -0.01)**	**-0.30**	**(-0.58 , -0.01)**
$\psi_{females}, I$:β_{jump}	**-0.76**	**(-0.98 , -0.54)**	**-0.76**	**(-0.98 , -0.54)**

the capture probability is significantly higher for male newts than female newts. This model also shows that the survival probability decreases when the mean minimum winter temperature increases; this is in agreement with the previous findings from Griffiths et al. (2010). We note that modelling the parameters as a function of environmental covariates removed the issue of boundary estimates since none of the parameters hit the boundary. The model with additive gender and time effects, and p_2 and p_6 fixed to 1 (denoted by L4 in Table 6.5), however, showed no significant differences between genders or sampling occasions in terms of capture probabilities. We also noticed that model $p(g), \phi(t), \psi\{(D+J)_{diff} * g\}(B)$ from level 4 necessitated fixing ϕ_3 to 1, which corresponds to the survival in year 2005. Survival in this year was also noted to behave differently in Griffiths et al. (2010), although it was unusually low in those analyses and unusually high in ours. Clearly, further investigation is warranted as to the modelling of capture and survival probabilities. However, in regards to our aim, there is no evidence of a weight-effect on either capture or survival probability.

Finally, we have insisted on the importance of goodness-of-fit throughout this thesis. At first sight, multi-state models seem advantageous in this aspect since they have a goodness-of-fit test suite. However, in a context where the states represent categories of a finely discretised continuous variable, these tests cannot be used due to the sparseness of data per state and sampling occasion. Therefore, we just present the results of the goodness-of-fit tests for the simple CJS model by gender in Table 6.6. These results were obtained using program U-CARE. There is no obvious violation of the CJS model assumptions, although transience could be investigated since the one-sided test for males is not far from significance. (Note also that the result for test 2.CL comes from only the 9th occasion, there was not enough data to perform it for any of the other occasions.)

Table 6.6: Newts, goodness-of-fit of the fully time-dependent CJS model, by group (gender)

Test	Males		Females	
	dof	p-value	dof	p-value
3.SR	8	0.79	9	0.99
3.Sm	6	0.80	3	1
2.CT	4	0.79	4	0.74
2.CL	NA	NA	1	0.07
Transience (one-sided)	-	0.09	-	0.30
Trap-dependence (two-sided)	-	0.51	-	0.34
Global		dof		p-value
		35		0.995

6.3.2 Meadow voles

We revisited the previous analyses of meadow voles by Nichols et al. (1992), who used a multi-state framework with coarse discretisation (only 4 categories) to model body mass. The dataset consisted of CJS-format information over four sampling occasions, for 515 voles captured at the Patuxent Wildlife Center in Laurel, Maryland, from fall 1981 to spring 1982 (Nichols et al., 1992). One record was deleted due to the fact that no mass was assigned at any of the capture occasions. As done previously, we used a finer discretisation with $m = 20$ to explore a potential effect of weight on survival or capture; this yielded a model with 18 states. We used the model structure and step-up approach described in Section 6.3.1, also using the same model notations, the effective sample size for this dataset was 819. The results of the step-up procedure are given in Table 6.7. We stopped the model fitting at level 3, given that there was no support for a model with weight-dependence for either capture or survival probability. Model $p(w), \phi(t), \psi(D+J)_{diff}$ has a $\Delta(AIC_c) = 2.7$ compared to the best model of level 3, but a likelihood-ratio test between models $p(w), \phi(t), \psi(D+J)_{diff}$ and $p(.), \phi(t), \psi(D+J)_{diff}$ indicated that the

weight effect was not significant (LRT-statistic = 1.96, p = 0.16 with d.o.f=1). The best model selected amongst our candidates models is $p(.), \phi(t), \psi\{(D + J)_{diff} + t\}$ with $\text{AIC}_c = 1563.86$.

The results of the best model from level 3 are given in Table 6.8. Since the survival and capture probabilities were modelled using a logit link, the standard errors on the probability scale were computed using the delta method described in Section 6.2.1. The survival probability was found to be significantly lower over the first period than the second period, which is consistent with the findings of Nichols et al. (1992). The capture probability was estimated to be 0.91, which is very close to the value of 0.9 reported by Bonner and Schwarz (2006) in their analysis of a subset of voles from the same dataset (Nichols et al. (1992) did not report the capture probability estimate). Once again, transition parameters are presented more for informative purposes. From Table 6.8, the log-odds of changing weight category relative to remaining in the same category is significantly lower in period 3 than period 1. The log-odds of increasing weight category relative to remaining in the same weight category significantly decreases with the departure state (jump size and occasion held constant) and the jump size (departure state and occasion held constant); whilst the log-odds of decreasing weight category relative to remaining in the same weight category, significantly decreases with the jump size.

In conclusion, using a multi-state approach with fine discretisation to model weight for the meadow voles, has shown no evidence of a weight-effect on either survival or capture. Finally, for informative purposes, we also present the goodness-of-fit results of the basic CJS model in Table 6.9; these tests were performed using program U-CARE. We note that the test for transience is at the limit of significance and could be investigated further, but that is not done here.

Table 6.7: Meadow voles, step-up model selection, the best model at each level is denoted in red

Model	Deviance	d	AIC	AIC_c
$p(.), \phi(.), \psi(.)$	2041.44	3	2047.44	2047.47
$p(.), \phi(.), \psi(D+J)_{diff}$	1561.83	8	1577.83	1578.01
$p(.), \phi(.), \psi(D+J)$	1714.63	5	1724.63	1724.70
$p(.), \phi(.), \psi(J)$	1748.97	4	1756.97	1757.02
$p(w), \phi(.), \psi(.)$	1920.71	4	1928.71	1928.76
$p(.), \phi(.), \psi(D)$	2008.52	4	2016.52	2016.57
$p(.), \phi(t), \psi(.)$	2025.81	5	2035.81	2035.88
$p(t), \phi(.), \psi(.)$	2035.42	5	2045.42	2045.49
$p(.), \phi(.), \psi(t)$	2035.95	5	2045.95	2046.02
$p(.), \phi(w), \psi(.)$	2041.14	4	2049.14	2049.18
$p(.), \phi(t), \psi(D+J)_{diff}$	1546.20	10	1566.20	1566.47
$p(.), \phi(.), \psi\{(D+J)_{diff}+t\}$	1555.11	10	1575.11	1575.38
$p(t), \phi(.), \psi(D+J)_{diff}$	1555.81	10	1575.81	1576.08
$p(t), \phi(.), \psi(D+J)_{diff}$	1555.81	10	1575.81	1576.08
$p(w), \phi(.), \psi(D+J)_{diff}$	1560.28	9	1578.28	1578.50
$p(.), \phi(w), \psi(D+J)_{diff}$	1561.45	9	1579.45	1579.68
$p(.), \phi(.), \psi\{(D+J)_{diff}*t\}$	1545.95	20	1585.95	1587.01
$p(.), \phi(t), \psi\{(D+J)_{diff}+t\}$	1539.48	12	1563.48	1563.86
$p(w), \phi(t), \psi(D+J)_{diff}$	1544.24	11	1566.24	1566.56
$p(t), \phi(t), \psi(D+J)_{diff}$	1543.87	12	1567.87	1568.26
$p(.), \phi(t+w), \psi(D+J)_{diff}$	1546.19	11	1568.19	1568.52
$p(.), \phi(t.w), \psi(D+J)_{diff}$	1545.90	13	1571.90	1572.35
$p(.), \phi(t), \psi\{(D+J)_{diff}*t\}$	1530.32	22	1574.32	1575.59

Table 6.8: Voles, best model results, the significant effects are denoted in bold. For the transition parameters, I and D respectively denote weight increase and weight decrease, r denotes the departure state.

Parameter	Estimate	95% CI
ϕ_1	**0.67**	**(0.59 , 0.75)**
ϕ_2	**0.88**	**(0.82 , 0.94)**
ϕ_3	0.81	(0.73 , 0.89)
p	0.91	(0.87 , 0.96)
ψ, D: intercept	-1.28	(-3.09 , 0.52)
ψ, D:β_r	0.13	(-0.01 , 0.26)
ψ, D:β_{jump}	**-0.93**	**(-1.18 ,-0.67)**
ψ, I: intercept	6.21	(4.55 , 7.86)
ψ, I: β_r	**-0.61**	**(-0.76 ,-0.46)**
ψ, I:β_{jump}	**-0.62**	**(-0.74 , -0.51)**
ψ: β_{t2}	-0.48	(-1.19 , 0.23)
ψ: β_{t3}	**-0.89**	**(-1.59 , -0.20)**

Table 6.9: Voles, goodness-of-fit tests for time-dependent CJS model

Test	dof	p-value
3.SR	2	0.19
3.Sm	1	1
2.CT	1	0.33
2.CL	NA	NA
Transience (one-sided)	-	0.07
Trap-dependence (two-sided)	-	0.33
Global	4	0.37

6.4 Some possible extensions

The methodology of using a multi-state framework with fine discretisation to model an individual time-varying covariate could be extended to cases where this covariate is subject to a measurement error or when the covariate value is missing even when the animal is resighted. We explored the performance of the multi-state approach for these scenarios, using the simulated datasets from section 6.2, with $p = 0.9$, with the following modifications:

- adding a measurement error following a normal distribution with mean 0 s.d.= 0.1
- adding a measurement error following a normal distribution with mean 0 s.d.= 0.5
- rendering 10% of the observations missing completely at random

The results are presented in Table 6.10, using a discretisation of $m = 20$. The multi-state approach performs well even with measurement error, and, as expected, for the larger measurement error, the relative bias is slightly higher for the survival parameters while the coverage probability is slightly lower. When there were missing covariate values, we slightly modified the model used. Indeed, we defined an additional state corresponding to "*alive with covariate value unknown*". Recall that the transition probabilities were modelled very simply in section 6.2 as a function of the weight jump and survival as a function of the state (i.e. weight). These relationships do not make sense for the additional state denoted U since it does not have the same meaning as the other states; it does not represent a specific weight category. Indeed, it is not a state per se, but rather a modelling tool. Hence, the survival probability in state U is defined simply as an average of the survival probabilities in all possible states, with R denoting the number of "known" states: $\phi_U = \frac{\sum_{r=1}^{R} \phi r}{R}$. As for the transition probabilities, they are assumed to be uniform across the

different states for moving from the *alive with covariate value unknown* to any of the other live states: $\psi(U, s) = \frac{1}{R+1}$, whilst the probability of going from a known covariate value to an unknown covariate value is modelled, using a multinomial logit link, as a function of an additional parameter β_7. Table 6.10 shows that this approach yields results close to those obtained in section 6.2.

Further research is necessary, for instance looking at lower capture probabilities, higher percentage of missing values and so forth. Nonetheless, the few scenarios explored for an individual time-varying covariate measured with error or presenting missing values show promising results for the multi-state approach with a fine discretisation.

6.5 Conclusion

In this chapter, we have shown that a multi-state approach with fine discretisation performs well to incorporate an individual time-varying covariate in a CJS-type capture-recapture analysis. The multi-state approach presents distinct advantages over the existing methods that handle the problematic individual time-varying covariates. They use all available data (unlike conditional tools such as the trinomial method), they do not require modelling the covariate process (unlike the Langrock and King (2013) or the Bonner and Schwarz (2006) approaches). Finally there is no need for assumptions on the pattern missing covariate values due to non-captures (unlike the Worthington et al. (2015) method for instance) since these are automatically treated in the multi-state framework. Hence, the next stage for future research, would be to consolidate our results by comparing the performance of this multi-state approach with the other methods, exploring in particular situations where these might be affected (e.g. completely erroneous covariate model). We also note that in practice, for our newt application from Section 6.3.1, the multi-state approach was computationally much faster to run that the HMM model. Al-

Table 6.10: Simulated datasets with added measurement error or with missing data (MI): results from a multi-state approach, using a discretization with $m = 20$, ME(s.d.) denotes the standard error of the normal distribution used for the added measurement error ; RB denotes the relative bias, q the quantiles, CIW the confidence interval widths, CP the coverage probability and N the number of usable datasets.

Scenario	m	Parameter	RB($q_{0.025}$, $q_{0.975}$)	CIW	CP	N
ME (0.1)		β_0	-0.01 (-0.14, 0.15)	1.47	0.94	250
		β_1	-0.01 (-0.11, 0.13)	0.18	0.96	
		p	0.00 (-0.02, 0.02)	0.04	0.94	
ME (0.5)		β_0	(-0.05, -0.17) 0.1	1.42	0.88	250
		β_1	-0.04 (-0.14, 0.09)	0.17	0.88	
		p	0.00 (-0.02, 0.02)	0.04	0.93	
MI	20	β_0	0.00 (-0.14, 0.16)	1.61	0.96	250
		β_1	0.01 (-0.10, 0.15)	0.20	0.97	
		p	0.00 (-0.02, 0.02)	0.04	0.92	

though the timing might be affected by the software used (Matlab versus R), the fact that the HMM likelihood is constructed from individual likelihoods whilst the multi-state likelihood is computed efficiently using matrix-products based on sufficient statistics surely plays a role in the speed differences. Thus, computational speed could be another advantage for the multi-state approach and this should be explored further.

Based on our simulations, a fine discretisation with $m = 20$ intervals presented good properties; but these simulations represented a specific covariate (weight) for a specific animal (great crested newt). In a more general setting, the choice of m will depend on the range of the individual time-varying covariate of interest, and its precision of measure. For instance, if weight is measured to the nearest gram, it would not be informative to use intervals smaller than a gram. The points raised by Bonner and Schwarz (2006) regarding the problems linked to having a large number of states such as estimation problems, or

exhaustion of data also favour the choice of a not too large m. In particular, parameter redundancy could be induced by some of the states never being observed, when fitting state-dependent parameters. On the other hand, m should not be chosen as too small in order to avoid losing information. Finally, m should be chosen so that the multi-state assumption of same properties for animals in the same state is biologically reasonable. Based on this, a good compromise for m could be chosen and the model selection performed using this m. Then, for the best model, the recommendation of Langrock and King (2013) could be applied: repeating the model fitting for increasing values of m in order to improve the estimates' precision and stopping when the values do not change anymore. Note that this step did not seem of particular interest for our applications from Section 6.3 because none of our parameters of interest (capture and survival probabilities) were significantly linked to weight.

In this chapter, we handled the potential estimation issues linked to the relatively large number of states (20), by using simple structures to model the different parameters: incorporating the state (thus the covariate) using a logistic regression link for capture and survival probabilities as is usually done for continuous covariates. Since the discretisation is fine, we do not expect to lose much information. In addition to this, we also use very basic structures to model the transition probabilities. In doing so, we lose the advantage of the multi-state approach with regard to estimating the parameters freely (Bonner and Schwarz, 2006). As we pointed out in the previous sections, this leads to the raw transition parameters not being intuitively interpretable, but this can be remedied by examining the transition probability matrix rather than the raw estimates. Also, the simple structure might not represent the most accurate modelling of the transition process so one should remain cautious as to the interpretation of the transition parameters. In particular, the choice of a multinomial logistic-type link implies proportional odds, which might not always be realistic. However, since the aim is the estimation of survival and

capture parameters, and that we are able to do so accurately, we deem the simple transition structure to be an acceptable compromise for the situations examined. It remains to be seen if this simple structure performs well in more general settings. Finally, although we did not encounter any obvious identifiability issues in our model fitting, it would be of interest to formally prove the identifiability of these models using the procedures from Chapter 5.

Based on all these considerations, at this stage of the work, we recommend using a multi-state model with fine discretisation to incorporate the effect of an individual time-varying covariate on capture or survival probability, if that is the main aim of the analysis, and that the covariate process itself is not of much interest. However, if one is interested in the covariate itself and the biological interpretation of its variation over time, it will be more relevant to use one of the alternative methods which include the modelling of the covariate process. Further work is warranted regarding the goodness-of-fit assessment of these multi-state models with a large number of states representing a continuous covariate, since the usual goodness-of-fit suite is unlikely to be applicable (for reasons stated in Section 6.3.1). The use of score-tests for example could be examined (although they might pose a problem for boundary estimates). Finally, the extensions to measurement error and missing values show promising results but need to be explored in more detail. Again, the performance of the multi-state approach could be compared with the other existing methods for these situations. We also note that using a multievent framework would allow these issues to be described more elegantly, with less restrictive assumptions. Indeed, if measurement error is expected, an observation in state r might be modelled as being possibly generated by states $r-1$, r or $r+1$ (or other adjacent states). As for missing values, a multi-event model, with observation "weight unknown" being generated by any of the live states, would allow the missing data pattern to be more general than just missing at random.

As a general aside, all the methods explored seem to deal mainly with weight as the individual time-varying covariate process and a possible logistic-type relationship between survival/capture and the covariate. Future work could also include examining other types of covariates, where there is potentially less knowledge about the covariate process, and exploring other relationships between the parameter and weight, in particular the effects of a misspecified model. For instance, one could think of cases where the relationship between weight and survival probability would be an inverted U-type curve, with underweight and overweight animals both presenting lower survival probabilities.

Chapter 7

Conclusion

7.1 Contributions

In this thesis, we have mainly developed new diagnostic tools for various capture-recapture models by focusing on specific violations of model assumptions. These tools lead to conclusions that allow us to guide the model-building process and give biological insights to the data.

In Chapter 2, we developed a new test for detecting heterogeneity in capture within a Cormack-Jolly-Seber (CJS) framework, based on Goodman-Kruskal's gamma. We used simulation to compare its performance to that of existing methods, which we described in detail: the existing CJS diagnostic tests, the Leslie test for equal catchability and Carothers' extension of this test. Simulation results have shown this test to be powerful and also less sensitive to trap-happiness than existing tests. If this test is significant, then models accounting for heterogeneity in capture should be considered and fitted.

Chapter 3 progressed to the more general multi-state models, for which we also employed Goodman-Kruskal's gamma, albeit in a different manner, to detect the existence of a mover-stayer structure. This phenomenon, and

more generally the violation of the assumption of homogeneous transitions for animals in the same state, has not received much attention, although it could provide interesting ecological insights. A simulation study allowed us not only to draw conclusions about the new test, but also revealed new findings regarding the interpretation of the existing diagnostic test WBWA, which we also described in detail. Indeed, the WBWA test is usually used as an indication of memory, but our simulation study found it to also be sensitive to heterogeneity in movement or preferences; whilst the new test was found to be sensitive to both memory and a mover-stayer structure. Adapting the WBWA test and combining this adaptation with a slightly modified version of the test of positive association allowed us to detect and differentiate between heterogeneity in movement or preferences, memory and a mover-stayer structure.

Chapter 4 focussed on the presence of partial observations in a multistate capture-recapture experiment. This situation, where the state cannot be determined upon capture for some of the animals, can be expressed as a special case of the multievent model. We theoretically showed that if partially observed animals actually belong to one of the observable states defined in the experiment, then those partially observed at a given occasion i and seen again at later occasions follow a multinomial distribution that is a mixture of the distributions followed by animals observed at i in the observable states and seen again at the same later occasions. Based on this mixture property, we derived a test for mixtures, following a path similar to the construction of existing Test M, which allows us to assess whether partial observations are generated by the observable states. Using simulations under very good conditions, the test was shown to work as expected theoretically.

In Chapter 5, we sought to derive a test for unobservable states. This led us through a journey which differed from the previous chapters. Indeed we expected a straightforward extension of the existing Test M, based on the fact that unobservable states translated into unobservable bases. Only the

test, which uses the Multinomial Maximum Likelihood Mixture method relies on two steps: parameter estimation and goodness-of-fit assessment; and the model with unobservable bases was found to be parameter redundant. Hence, the concept of parameter redundancy was introduced and examined in this framework. This led to deriving general parameter redundancy results in a context of multinomial mixtures and bases. Afterwards, we used simulation, focussing on one and two unobservable bases, under very good conditions to verify that the test for unobservable bases worked as expected. In doing this, we realised that in the reparametrised model that used some of the mixtures as proxy bases for the unobservable ones, the remaining original mixtures were not necessarily mixtures, but *linear combinations* of the new bases (proxy and original). We also had to use an iterative optimisation method to avoid getting stuck in local optima. Finally, we derived a step-up procedure based on tests for linear combinations, using the reparametrisations resulting from the parameter redundancy work, that indicated whether there was statistical support for the existence of one or two unobservable bases. The procedure worked as expected in theory on the simulations considered. It was only after all these investigations that we came back to the specific capture-recapture setting with unobservable states, which constituted one possible area of application of our results. Simulations under very large sample sizes showed that the procedure can work to detect one or two unobservable states in a capture-recapture setting.

Finally, Chapter 6 differed from the rest of the thesis in its scope. In this chapter we used multi-state models with a fine discretisation, imposing simple structures on the model parameters, to incorporate the effect of individual time-varying covariates such as body mass on the model parameters. This approach showed promising results as well as a potential to be extended to more general situations such as covariates measured with error or the presence of missing data.

Note that, by essence, biological phenomena are very complex and this complexity will not be captured in its entirety by the various model assumption violations considered. However, models, by nature, are themselves a simplification of reality and the diagnostic tools developed throughout this thesis provide guidance as to directions to explore, both from a biological and modelling perspective; they can also help avoid drawing conclusions based on erroneous assumptions.

7.2 Future work

Building on the conclusions of this thesis, three main areas of future research can be identified: the testing procedure itself, sample sizes and additional ecological applications. Firstly, regarding the testing process, mixture and/or linear combination tests could potentially be extended for more general multievent models, when there is uncertainty on all observations, to test for the number of underlying states necessary to provide an adequate fit to the data. Also, for the individual time-varying covariates, re-framing the problem in a multi-state context could allow for the use of score tests combined with a step-up model building approach for goodness-of-fit assessment and this should be explored further.

Secondly, we used simulations under very good conditions i.e. when sample sizes are large, to check that the new diagnostic tools worked as expected in theory and reacted to specific phenomena. It would be of interest to perform power analyses and assess the test properties for more moderate sample sizes, especially for the tools from Chapters 4 and 5. The positive association tests from Chapters 2 and 3 have been evaluated on more moderate sample sizes, however one could look into adapting these tools or potentially deriving new ones for small sample sizes.

Finally, the methodologies developed throughout this thesis have generally been applied to real datasets, and mostly gave the results we expected: Great cormorants and Sandwich terns (Chapter 2), Canada Geese (Chapters 3 to 5), Greater flamingoes (Chapter 4), Great crested newts and Meadow voles (Chapter 6). However, they would benefit from exploring a wider variety of ecological applications, delving into the effects of fitting inadequate models as well as examining new motivating datasets that would benefit from the insights provided by our tools.

Bibliography

Abadi, F., Botha, A., and Altwegg, R. (2013). Revisiting the effect of capture heterogeneity on survival estimates in capture-mark-recapture studies: Does it matter? *PLoS ONE*, 8(4):e62636. Available at `https://doi.org/10.1371/journal.pone.0062636`.

Amstrup, S. C., McDonald, T. L., and Manly, B. F. J. (2005). *Handbook of capture-recapture analysis*. Princeton University Press, Princeton, New Jersey, USA.

Bonner, S. J., Morgan, B. J. T., and King, R. (2010). Continuous covariates in mark-recapture-recovery analysis: a comparison of methods. *Biometrics*, 66(4):1256–1265.

Bonner, S. J. and Schwarz, C. J. (2006). An extension of the Cormack–Jolly–Seber model for continuous covariates with application to Microtus pennsylvanicus. *Biometrics*, 62(1):142–149.

Boos, D. D. and Stefanski, L. A. (2011). P-value precision and reproducibility. *The American Statistician*, 65(4):213–221.

Box, G. E. P. and Draper, N. R. (1987). *Empirical model-building and response surfaces*. Wiley, New York, USA.

Brown, M. B. and Benedetti, J. K. (1977). Sampling behavior of tests for cor-

relation in two-way contingency tables. *Journal of the American Statistical Association*, 72(358):309–315.

Burnham, K. P. (1991). On a unified theory for release-resampling of animal populations. In Chao, M. T. and Cheng, P. E., editors, *Proceedings of the 1990 Taipei Symposium in Statistics*, pages 1–35. Institute of Statistical Science, Academia Sinica. Taipei, Taiwan, R.O.C.

Burnham, K. P. and Anderson, D. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2):261–304.

Burnham, K. P. and Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, New York, USA.

Burnham, K. P., Anderson, D. R., White, G. C., Brownie, C., and Pollock, K. H. (1987). Design and analysis methods for fish survival experiments based on release-recapture. *American Fisheries Society Monograph*, 5:1–437.

Carothers, A. D. (1971). An examination and extension of Leslie's test of equal catchability. *Biometrics*, 27(3):615–630.

Catchpole, E. A. and Morgan, B. J. T. (1997). Detecting parameter redundancy. *Biometrika*, 84(1):187–196.

Catchpole, E. A., Morgan, B. J. T., and Freeman, S. N. (1998). Estimation in parameter-redundant models. *Biometrika*, pages 462–468.

Catchpole, E. A., Morgan, B. J. T., and Tavecchia, G. (2008). A new method for analysing discrete life history data with missing covariate values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(2):445–460.

Choquet, R., Lebreton, J.-D., Gimenez, O., Reboulet, A.-M., and Pradel, R. (2009a). U-CARE: Utilities for performing goodness of fit tests and manipulating CApture-REcapture data. *Ecography*, 32(6):1071–1074.

Choquet, R. and Nogue, E. (2006). E-surge 1.8 user's manual. *CEFE, Montpellier, France*. Available online at `https://www.cefe.cnrs.fr/images/stories/DPTEEvolution/biostatistiques/LOGICIELS/E-SURGE-MANUAL.pdf`.

Choquet, R., Reboulet, A.-M., Lebreton, J.-D., Gimenez, O., and Pradel, R. (2005). U-CARE 2.2 user's manual. *CEFE, Montpellier, France*. Available online at `https://www.cefe.cnrs.fr/images/stories/DPTEEvolution/biostatistiques/Chercheurs/remi_choquet/Choquet-USER%20MANUAL%20U-CARE%202.2.pdf`.

Choquet, R., Rouan, L., and Pradel, R. (2009b). Program E-SURGE: a software application for fitting multievent models. In Thomson, D. L., Cooch, E. G., and Conroy, M. J., editors, *Modeling demographic processes in marked populations*, volume 3 of *Environmental and Ecological Statistics*, pages 847–868. Springer, New York, USA.

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, 37(3/4):256–266.

Cole, D. J. (2012). Determining parameter redundancy of multi-state mark–recapture models for sea birds. *Journal of Ornithology*, 152(2):305–315.

Cole, D. J., Morgan, B. J. T., Catchpole, E. A., and Hubbard, B. A. (2012). Parameter redundancy in mark-recovery models. *Biometrical Journal*, 54(4):507–523.

Cole, D. J., Morgan, B. J. T., McCrea, R. S., Pradel, R., Gimenez, O., and Choquet, R. (2014). Does your species have memory? analyzing capture–

recapture data with memory models. *Ecology and Evolution*, 4(11):2124–2133.

Cole, D. J., Morgan, B. J. T., and Titterington, D. M. (2010). Determining the parametric structure of models. *Mathematical biosciences*, 228(1):16–30.

Conn, P. B. and Cooch, E. G. (2009). Multistate capture–recapture analysis under imperfect state observation: an application to disease models. *Journal of Applied Ecology*, 46(2):486–492.

Conover, W. (1980). *Practical Nonparametric Statistics*. John Wiley & Sons, New-York, USA.

Cooch, E. and White, G. (2014). Program MARK: a gentle introduction. Available online at `http://www.phidot.org/software/mark/docs/book/`.

Corkrey, R., Brooks, S., Lusseau, D., Parsons, K., Durban, J. W., Hammond, P. S., and Thompson, P. M. (2012). A Bayesian capture–recapture population model with simultaneous estimation of heterogeneity. *Journal of the American Statistical Association*, 103(483):948–960.

Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society. Series B (Methodological)*, 46(3):440–464.

Cressie, N. and Read, T. R. C. (1988). *Goodness-of-fit statistics for discrete multivariate data*. Springer, New York, USA.

Cubaynes, S., Lavergne, C., Marboutin, E., and Gimenez, O. (2012). Assessing individual heterogeneity using model selection criteria: how many mixture components in capture-recapture models? *Methods in Ecology and Evolution*, 3(3):564–573.

Cubaynes, S., Pradel, R., Choquet, R., Duchamp, C., Gaillard, J.-M., Lebreton, J.-D., Marboutin, E., Miquel, C., Reboulet, A., Poillot, C., Taberlet, P.,

and Gimenez, O. (2010). Importance of accounting for detection heterogeneity when estimating abundance: the case of French wolves. *Conservation Biology*, 24(2):621–626.

Everitt, B. and Hand, D. (1981). *Finite Mixture Distributions, Monograph on applied Probability and Statistics*. Chapman and Hall, Ltd, Boca Raton, USA.

Everitt, B. S. (1992). *The Analysis of contingency tables*. Chapman & Hall, London, UK.

Fletcher, D., Lebreton, J.-D., Marescot, L., Schaub, M., Gimenez, O., Dawson, S., and Slooten, E. (2012). Bias in estimation of adult survival and asymptotic population growth rate caused by undetected capture heterogeneity. *Methods in Ecology and Evolution*, 3(1):206–216.

Genovart, M., Pradel, R., and Oro, D. (2012). Exploiting uncertain ecological fieldwork data with multi-event capture–recapture modelling: an example with bird sex assignment. *Journal of animal ecology*, 81(5):970–977.

Griffiths, R. A., Sewell, D., and McCrea, R. S. (2010). Dynamics of a declining amphibian metapopulation: survival, dispersal and the impact of climate. *Biological Conservation*, 143(2):485–491.

Halsey, L. G., Curran-Everett, D., Vowler, S. L., and Drummond, G. B. (2015). The fickle p value generates irreproducible results. *Nature methods*, 12(3):179–185.

Hammond, P. S., Mizroch, S. A., and Donovan, G. P. (1990). *Report of the International Whaling Commission (Special Issue 12). Individual Recognition of Cetaceans: Use of Photo-Identification and Other Techniques to Estimate Population Parameters.* International Whaling Commission, Cambridge, UK.

Hénaux, V., Bregnballe, T., and Lebreton, J.-D. (2007). Dispersal and recruitment during population growth in a colonial bird, the great cormorant Phalacrocorax carbo sinensis. *Journal of Avian Biology*, 38:44–57.

Hestbeck, J. B., Nichols, J. D., and Malecki, R. A. (1991). Estimates of movement and site fidelity using mark-resight data of wintering Canada geese. *Ecology*, 72(2):523–533.

Kendall, W. L. (2004). Coping with unobservable and mis–classified states in capture–recapture studies. *Animal biodiversity and Conservation*, 27(1):97–107.

Kéry, M. and Schaub, M. (2012). Chapter 9 - estimation of survival and movement from capture-recapture data using multistate models. In Kéry, M., , and Schaub, M., editors, *Bayesian Population Analysis using WinBUGS*, pages 263 – 313. Academic Press, Boston, USA. Available online at `http://www.sciencedirect.com/science/article/pii/B9780123870209000092`.

King, R. and McCrea, R. S. (2014). A generalised likelihood framework for partially observed capture–recapture–recovery models. *Statistical Methodology*, 17:30–45.

Lachish, S., Knowles, S. C., Alves, R., Wood, M. J., and Sheldon, B. C. (2011). Infection dynamics of endemic malaria in a wild bird population: parasite species-dependent drivers of spatial and temporal variation in transmission rates. *Journal of Animal Ecology*, 80(6):1207–1216.

Langrock, R. and King, R. (2013). Maximum likelihood estimation of mark-recapture-recovery models in the presence of continuous covariates. *Annals of Applied Statistics*, 7(3):1709–1732.

Lebreton, J.-D., Burnham, K. P., Clobert, J., and Anderson, D. R. (1992).

Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs*, 62(1):67–118.

Lebreton, J.-D., Nichols, J. D., Barker, R. J., Pradel, R., and Spendelow, J. A. (2009). Modeling individual animal histories with multistate capture-recapture models. *Advances in Ecological Research*, 41:87–173.

Lebreton, J.-D. and Pradel, R. (2002). Multistate recapture models: Modelling incomplete individual histories. *Journal of Applied Statistics*, 29(1-4):353–369.

Lecoutre, J.-P. (2006). *Statistique et probabilités-3e édition*. Dunod, Paris, FR.

Madsen, J., Tjørnløv, R. S., Frederiksen, M., Mitchell, C., and Sigfusson, A. T. (2014). Connectivity between flyway populations of waterbirds: assessment of rates of exchange, their causes and consequences. *Journal of Applied Ecology*, 51(1):183–193.

Mastro, T. D., Kitayaporn, D., Weniger, B. G., Vanichseni, S., Laosunthorn, V., Uneklabh, T., Uneklabh, C., Choopanya, K., and Limpakarnjanarat, K. (1994). Estimating the number of HIV-infected injection drug users in Bangkok: a capture-recapture method. *American Journal of Public Health*, 84(7):1094–1099.

McCrea, R. S. and Morgan, B. J. T. (2011). Multistate mark-recapture model selection using score tests. *Biometrics*, 67(1):234–241.

McCrea, R. S. and Morgan, B. J. T. (2014). *Analysis of Capture-Recapture Data*. Chapman and Hall/CRC, Boca Raton, USA.

Meyer, D., Zeileis, A., and Hornik, K. (2016). vcd: Visualizing categorical data. R package version 1.4-3. Available online at `https://cran.r-project.org/web/packages/vcd/vcd.pdf`.

Moore, D. (1986). Tests of chi-squared type. In D'Agostino, R. and Stephens, M., editors, *Goodness-of-fit techniques*, volume 68, chapter 3, pages 63–96. CRC Press, New-York (USA).

Newman, K., Buckland, S. T., Morgan, B. J. T., King, R., Borchers, D. L., Cole, D. J., Besbeas, P., Gimenez, O., and Thomas, L. (2014). *Modelling population dynamics*. Springer, New York, USA.

Nichols, J. D., Sauer, J. R., Pollock, K. H., and Hestbeck, J. B. (1992). Estimating transition probabilities for stage-based population projection matrices using capture-recapture data. *Ecology*, 73(1):306–312.

North, P. M. and Morgan, B. J. T. (1979). Modelling heron survival using weather data. *Biometrics*, 35(3):667–681.

Oliver, L. J., Morgan, B. J. T., Durant, S. M., and Pettorelli, N. (2011). Individual heterogeneity in recapture probability and survival estimates in cheetah. *Ecological Modelling*, 222(3):776–784.

Orians, G. H. and Leslie, P. H. (1958). A capture-recapture analysis of a shearwater population: with a statistical appendix. *The Journal of Animal Ecology*, 27(1):71–86.

Péron, G., Crochet, P.-A., Choquet, R., Pradel, R., Lebreton, J.-D., and Gimenez, O. (2010). Capture-recapture models with heterogeneity to study survival senescence in the wild. *Oikos*, 119(3):524–532.

Pledger, S., Pollock, K. H., and Norris, J. L. (2003). Open capture-recapture models with heterogeneity: I. Cormack-Jolly-Seber model. *Biometrics*, 59(4):786–794.

Pollock, K. H. (2002). The use of auxiliary variables in capture-recapture modelling: an overview. *Journal of Applied Statistics*, 29:85–102.

Pradel, R. (2005). Multievent: An extension of multistate capture-recapture models to uncertain states. *Biometrics*, 61(2):442–447.

Pradel, R., Gimenez, O., and Lebreton, J.-D. (2005). Principles and interest of GOF tests for multistate capture-recapture models. *Animal Biodiversity and Conservation*, 28(2):189–204.

Pradel, R., Wintrebert, C. M. A., and Gimenez, O. (2003). A proposal for a goodness-of-fit test to the Arnason-Schwarz multisite capture-recapture model. *Biometrics*, 59(1):43–53.

Prévot-Julliard, A. C., Lebreton, J. D., and Pradel, R. (1998). Re-evaluation of adult survival of Black-headed Gulls (Larus ridibundus) in presence of recapture heterogeneity. *The Auk*, 115(1):85–95.

Reise, S. P. and Revicki, D. A. (2014). *Handbook of item response theory modeling: Applications to typical performance assessment*. Routledge, New-York (USA).

Romano, J. P., Shaikh, A. M., and Wolf, M. (2010). Multiple testing. In Durlauf, S. N. and Blume, L. E., editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, Basingstoke, UK.

Rouan, L., Choquet, R., and Pradel, R. (2009). A general framework for modeling memory in capture-recapture data. *Journal of Agricultural, Biological, and Environmental Statistics*, 14(3):338–355.

Sanz-Aguilar, A., Bechet, A., Germain, C., Johnson, A. R., and Pradel, R. (2012). To leave or not to leave: survival trade-offs between different migratory strategies in the greater flamingo. *Journal of Animal Ecology*, 81(6):1171–1182.

Siegel, S. and Castellan Jr., N. J. (1988). *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill Book Company, New York, USA.

Sokal, R. R. and Rohlf, F. J. (2012). *Biometry (4th edn)*. WH Freeman and Company, New York, USA.

Spilerman, S. (1972). Extensions of the mover-stayer model. *American Journal of Sociology*, 78(3):599–626.

Upton, G. and Cook, I. (2008). Fisher exact test. In *A Dictionary of Statistics*. Oxford University Press. Available at `http://www.oxfordreference.com.chain.kent.ac.uk/view/10.1093/acref/9780199541454.001.0001/acref-9780199541454-e-608`.

van der Heijden, P. G. M., de Vries, I., Böhning, D., and Cruyff, M. (2015). Estimating the size of hard-to-reach populations using capture-recapture methodology, with a discussion of the international labour organization's global estimate of forced labour. In *Forum on Crime and Society, Special issue: Researching hidden populations: approaches to and methodologies for generating data on trafficking in persons.*, volume 8, page 109, United Nations, New York, USA.

Williams, B. K., Nichols, J. D., and Conroy, M. J. (2002). *Analysis and Management of Animal Populations*. Academic Press, San Diego, California, USA.

Worthington, H., King, R., and Buckland, S. T. (2015). Analysing mark–recapture–recovery data in the presence of missing covariate data via multiple imputation. *Journal of Agricultural, Biological, and Environmental Statistics*, 20(1):28–46.

Yantis, S., Meyer, D. E., and Smith, J. K. (1991). Analyses of multinomial mixture distributions: new tests for stochastic models of cognition and action. *Psychological Bulletin*, 110(2):350.